AD-A100 319 MINNESOTA UNIV ST PAUL DEPT OF ELECTRICAL ENGINEERING F/G 9/3
ADAPTIVE HIGH RESOLUTION SPECTRAL ANALYSIS WITH NOISY DATA.(U)
DEC 80 M KAVEH AFOSR-78-3628

UNCLASSIFIED AFOSR-TR-81-0489 NL

1 OF 1
AD
A100319

END
DATE
FILMED
7 81
DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-81-0489
2. GOVT ACCESSION NO.: AD-A100 319
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): ADAPTIVE HIGH RESOLUTION SPECTRAL ANALYSIS WITH NOISY DATA
5. TYPE OF REPORT & PERIOD COVERED: FINAL 16 JUN 78-15 DEC 80
6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): M. Kaveh
8. CONTRACT OR GRANT NUMBER(s): AFOSR-78-3628

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Electrical Engineering
University of Minnesota
St. Paul MN 55114

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: PE61102F 2304/A6

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research/NM
Bolling AFB DC 20332

12. REPORT DATE: DEC 1980
13. NUMBER OF PAGES: 79

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes the results of research during the past two years, in understanding the resolution of the autoregressive (AR) spectral estimators and developing and evaluating computationally efficient autoregressive-moving average (ARMA) spectral estimators. The loss in the resolution of the AR spectral estimator in the presence of noise is related to the appearance of zeros in the z-plane. A parallel resonator model is

(CONT.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A100319

DTIC FILE COPY

(ITEM #20, CONT.)

developed to relate the loss in resolution (bandwidth expansion) to the signal-to-noise ratio and parameters of the noiseless signal model.

A new technique for the identification of the order of an AR model was derived that shows substantial stability compared to the popular Akaike Information Criterion method. Order determination was emphasized, since increase in the order of the AR spectral estimator, to account for the presence of noise, is naturally accompanied by larger variance of the estimates and appearance of spurious peaks.

Several sub-optimum (non-maximum-likelihood) ARMA spectral estimators were also developed. These methods are computationally efficient, but statistically not very stable for small data records. An evaluation of the statistical properties of the different sub-optimum ARMA techniques led to the evaluation of asymptotic bounds on the variances of the estimates of the parameters or the poles and zeros of the model through the evaluation of Fisher's information matrix. Finally, a modification of Burg's MEM spectral estimator was developed that improves the accuracy of the spectral estimates of complex sinusoids and makes the method considerably more robust.

AFOSR-TR- 81-0489

FINAL REPORT
on
ADAPTIVE HIGH RESOLUTION SPECTRAL
ANALYSIS WITH NOISY DATA
GRANT: AFOSR-78-3628
PERIOD: June 16, 1978 -
December 15, 1980

PRINCIPAL INVESTIGATOR:
M. Kaveh

81 6 12 137

Approved for public release;
distribution unlimited.

SUMMARY OF ACTIVITIES

Dist
A

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

RESEARCH OBJECTIVES

1. Determine the non-asymptotic resolution of the autoregressive (AR) spectral estimator.
2. Relate the loss of resolution of the AR spectral estimator to the noise and interference present in the measured signal.
3. Develop computationally efficient autoregressive-moving average (ARMA) spectral estimators to obtain improved spectral resolution for noisy signals.
4. Investigate the properties of the ARMA spectral estimators, theoretically and by numerical simulations.
5. Develop a robust order determination scheme.

RELEVANT PUBLICATIONS

Published

1. M. Kaveh, "High resolution spectral estimation for noisy signals", IEEE Trans. on ASSP, June 1979.
2. M. Kaveh, "A modified Akaike information criterion, "Proceedings of IEEE Conf. on Decision and Control, December 1978.
3. S. Bruzzone and M. Kaveh, "On some suboptimum ARMA spectral estimators", IEEE Trans. on ASSP, December 1980.
4. G. Lippert and M. Kaveh, "Frequency errors in the "tapered" Burg spectral estimates of complex sinusoids", Proceedings of ICASSP '81.

Under Preparation

1) Statistical properties of optimum and suboptimum ARMA spectral estimator. For IEEE Trans. on Info. Theory.
2) An optimum taper for MEM spectral analysis. For IEEE Trans. on ASSP.
3) An order determination method for all-pole models. For IEEE Trans. on ASSP.

INTERACTIONS

I. Conferences

1. M. Kaveh, "High resolution spectral estimation via rational models", (Invited), first RADC Workshop on Spectrum Estimation, May 1978.
2. M. Kaveh and S. Bruzzone, "Order determination for autoregressive spectral estimation", (Invited), second RADC Workshop on Spectrum Estimation, Oct. 1979.
3. S. Bruzzone and M. Kaveh, "A Least-Squares Estimator for Unknown Signals Buried in AR Noise and its Asymptotic Properties", IEEE International Symposium on Information Theory, June 1979.
4. M. Kaveh and S. Bruzzone, "A computationally efficient suboptimum ARMA spectral estimator", IEEE Conference on Decision and Control, 1979.
5. M. Kaveh, "A comparative view of ARMA spectral estimation", (Invited), to be presented at the first ASSP Workshop on spectral estimation.

II. Relevant Consulting Activities

M. Kaveh has been consulting with Dr. James Evans of the Advanced Techniques Group of the M.I.T. Lincoln Laboratory, Lexington, MA, on "High resolution sampled aperture processing", since March 1979.

PROFESSIONAL PERSONNEL

1. M. Kaveh, Associate Professor of Electrical Engineering, PI.
2. S. P. Bruzzone, Research Assistant

   M.S. December 1978 (Partially supported by this grant).

   Thesis title: "Stochastic Modeling with Applications to Signal Detection and Estimation."

   Currently pursuing the Ph.D. degree, expected completion date: June 1981 (Supported by this grant).

   Thesis title: "Analysis of Some Suboptimum ARMA Spectral Estimators."
3. G. Lippert (Computer work supported by this grant).

   M.S. degree completed, June 1980.

   Thesis title: "Data-adaptive Array Processing."

TECHNICAL REPORT

TABLE OF CONTENTS

ABSTRACT

CHAPTER		PAGE
I.	INTRODUCTION	1
	1. Report Outline	2
II.	AUTOREGRESSIVE SPECTRAL ESTIMATION OF NOISY SIGNALS	4
	1. The AR Spectral Estimator	4
	2. Spectral Estimation in the Presence of Noise and Interference	6
	3. Parallel Resonator Model	10
	4. Analysis of a Single Resonator	13
	5. Perturbations of Pole Positions	17
III.	ORDER DETERMINATION FOR AR MODELS	25
	1. Akaike's Information Criterion	26
	2. The New Criterion	27
	3. Simulation Results	31
IV.	SOME SUBOPTIMUM ARMA SPECTRAL ESTIMATORS	34
	1. The Spectrum of an ARMA Process	35
	2. The LS Estimator	36
	3. The MYW Spectral Estimator	38
	4. Simulation Results	40
V.	STATISTICAL CLASSIFICATION OF SOME ARMA SPECTRAL ESTIMATORS	44
	1. Classification of Some ARMA Spectral Estimators	45
	2. Evaluation of Fisher's Information Matrix	48

TABLE OF CONTENTS (continued)

CHAPTER		PAGE
VI.	SPECTRAL ESTIMATION FOR NOISY COMPLEX SINUSOIDS	58
	1. Generalized Error Expression	59
	2. An Optimum Taper	63
	3. Simulation Results	64
VII.	SUMMARY AND CONCLUSIONS	68
APPENDIX A		70
APPENDIX B		74
REFERENCES		77

ABSTRACT

This report summarizes the results of research during the past two years, in understanding the resolution of the autoregressive (AR) spectral estimators and developing and evaluating computationally efficient autoregressive-moving average (ARMA) spectral estimators. The loss in the resolution of the AR spectral estimator in the presence of noise is related to the appearance of zeros in the z-plane. A parallel resonator model is developed to relate the loss in resolution (bandwidth expansion) to the signal-to-noise ratio and parameters of the noiseless signal model.

A new technique for the identification of the order of an AR model was derived that shows substantial stability compared to the popular Akaike Information Criterion method. Order determination was emphasized, since increase in the order of the AR spectral estimator, to account for the presence of noise, is naturally accompanied by larger variance of the estimates and appearance of spurious peaks.

Several sub-optimum (non-maximum-likelihood) ARMA spectral estimators were also developed. These methods are computationally efficient, but statistically not very stable for small data records. An evaluation of the statistical properties of the different sub-optimum ARMA techniques led to the evaluation of asymptotic bounds on the variances of the estimates of the parameters or the poles and zeros of the model through the evaluation of Fisher's information matrix. Finally, a modification of Burg's MEM spectral estimator was developed that improves the accuracy of the spectral estimates of complex sinusoids and makes the method considerably more robust.

I. INTRODUCTION

The problem of spectral estimation via the autoregressive (AR) and autoregressive-moving-average (ARMA) modeling of the observed noisy signal is treated in this report. High resolution spectral analysis based on the AR model has received much attention in recent years [1] - [3]. The basic impetus for taking these approaches is due to the fact that the parametric modeling schemes are data adaptive and are free from the effects of window functions that are inherent in the traditional Blackman and Tukey [4] type spectral estimators. Furthermore, the AR model is easily estimated making it useful in applications such as radar signal processing that require near-real-time processing.

The properties of the AR spectral estimator have been studied, theoretically in the asymptotic case [3], [5], [6] and empirically [1] , [2]. It has been shown that this estimator in many cases offers considerably higher resolution based on the same amount of data, than the Blackman and Tukey type estimators. Furthermore, the above asymptotic and empirical investigations have shown the variance of the AR estimates to be comparable to the unsmoothed Blackman and Tukey type estimates, for the same number of autocorrelation lags. It should be pointed out, however, that the AR estimates usually require much fewer lags for the same resolution.

Most practical applications of spectral analysis involve noisy signals and/or multiple signal and noise mixtures. Therefore, it is natural to consider the performance of the AR spectral estimator in the presence of noise. Previous studies

[1], [2] had shown resolution degradation in the presence of noise. The solution presented for improving resolution was given as increasing the order of the estimator, in a rather arbitrary fashion.

This report deals with the question of AR signals in noise. The main thrust of the work reported here was to improve spectral resolution by using ARMA models for the measured signals. Specifically some suboptimum but computationally efficient ARMA spectral estimation algorithms were developed and their properties studied. In the process of the investigations, an order determination scheme as well as a new Burg type spectral estimator, however, were also developed and will be described here.

I.1 Report Outline

This report is organized as follows. First the AR process is defined. The sum of uncorrelated AR processes and white noise is then considered and shown to be represented by an equivalent ARMA process. A multiple resonator model for signals of interest is presented and shown to be equivalent to the ARMA model. Resolution degradation as a function of noise and resonator pole locations is considered and from it some representative curves of bandwidth expansion are presented.

Chapter III. of the report deals with the basic question of AR model order determination. A new criterion related to that of Akaike's (AIC) is derived and comparative examples are given. The next two chapters treat the basic emphasis of this work, namely that of ARMA spectral estimation. First, several ARMA schemes are presented that are computationally efficient but sub-optimum. Sub-optimality is treated with respect to the maximum-likelihood (ML) estimates of the parameters. The subsequent

chapter then discusses some statistical properties of classes of ARMA spectral-estimator, with respect to the statistic used in the estimation. Here, again, the reference scheme is the ML estimator. The final chapter of the report deals with a new scheme, related to Burg's MEM method, which is especially useful for the common case of sinusoidal signals in noise. This method is shown to be more robust than the MEM technique, with better accuracy and equivalent or superior resolution.

II. AUTOREGRESSIVE SPECTRAL ESTIMATION OF NOISY SIGNALS

The most popular recent data-adaptive spectral estimation method is one based on an all-pole model for the signal. The algorithmic approach for the estimation of the parameters of such a system include, the methods of fitting the autoregressive (AR) coefficients to the data as well as the popular Burg maximum entropy method MEM. Because of the popularity of this model and computational simplicity of its estimates, the noisy signal spectral estimation will be confined to autoregressive signal models. That is, it will be assumed that in the absence of noise and interference an AR model satisfactorily describes the signal. Resolution degradation of the AR spectral estimates in the presence of noise and interference is then investigated and related to the changes in the model structure.

II.1 The AR Spectral Estimator

A zero-mean time series $\{x_t\}$ is said to satisfy an Lth order autoregressive model if:

$$x_t = \sum_{i=1}^{L} \alpha_i \, x_{t-i} + u_t \qquad \text{(II.1)}$$

where $\{\alpha_i\}$ denote the AR coefficients and $\{u_t\}$ is a zero-mean uncorrelated (white) sequence. Another interpretation of the model in (II.1) is that $\{\alpha_i\}$ represent an Lth order one-step ahead linear predictor of $\{x_t\}$. If $\{\alpha_i\}$ are then estimated, based on a minimum mean square error criterion $\{u_t\}$ on the average becomes an orthogonal sequence. It can be shown [2] that the model in (II.1) leads to a spectral density of the form

$$S_L(f) = \frac{S_1}{\left|1 - \sum_{k=1}^{L} \alpha_i e^{i2\pi k f \Delta T}\right|^2}, \quad |f| \le \frac{1}{2\Delta T} \tag{II.2}$$

where S_1 is the spectral level of the $\{u_t\}$ sequence.

The spectral density shown in (II.2) is the AR spectrum and it is this model that is fitted to an observed time series, by simply estimating $\{\alpha_i\}$ from the time series. Several estimation procedures for $\{\alpha_i\}$ have been discussed in the literature such as the maximum likelihood, the least-squares [7] and Burg's method based on forward and backward prediction error filtering [8]. If the number of data samples is not very small, the simplest and computationally most efficient estimates of $\{a_i\}$ are the solution of the Yule-Walker equations. These equations arise, simply by multiplying equation (II.1) by x_{t-i}, $i=1,\ldots L$ and taking the expectation of both sides, to obtain a relation between the autocorrelation function of the process $\{x_t\}$ and the coefficients $\{\alpha_i\}$. The Yule-Walker equations are then given by:

$$R_o A = \rho_o \tag{II.3}$$

where

$$R_o = \begin{vmatrix} r_o & r_1 & \cdots & r_{L-1} \\ r_1 & & & \\ \vdots & & & \\ r_{L-1} & & & r_o \end{vmatrix}, \quad A = \begin{vmatrix} \alpha_1 \\ \vdots \\ \alpha_L \end{vmatrix} \quad \text{and} \quad \rho_o = \begin{vmatrix} r_1 \\ \vdots \\ r_L \end{vmatrix}$$

and where r_i is the autocorrelation function of $\{x_t\}$ at the ith lag. Furthermore, the power in $\{u_t\}$ can be found by multiplying (II.1) by x_t and taking the expectation as:

$$S_1 = [r_o - \sum_{i=1}^{L} \alpha_i r_i]\Delta T \text{ , } \Delta T \text{ the sampling interval} \qquad \text{(II.4)}$$

In practice $\{\alpha_i\}$ and S_1 are estimated from (II.3) and (II.4) based on estimates of the autocorrelation function $\{\hat{r}_i\}$.

II.2 Spectral Estimation in the Presence of Noise and Interference

We now assume that the signal $\{x_t\}$ satisfies an Lth order AR model and therefore its spectrum can be estimated as in the previous section. The problem of interest is the estimation of the spectrum of the observed signal

$$y_t = x_t + \omega_t + n_t \qquad \text{(II.5)}$$

where $\{\omega_t\}$ is an AR(M) process and considered to be the interference and $\{n_t\}$ is a white noise sequence, with $\{n_t\}$, $\{x_t\}$ and $\{\omega_t\}$ mutually uncorrelated.

The resolution of the spectral estimators are now discussed in the asymptotic case, that is, when the autocorrelation function of $\{y_t\}$ is accurately known.

Let $\{x_t\}$ be described by (II.1) and $\{\omega_t\}$ be given by the following autoregressive model

$$\omega_t = \sum_{i=1}^{M} b_i \, \omega_{t-i} + v_t$$

The z-spectrum of y_t is then given by:

$$S_Y(z) = \frac{S_1}{D_x(z)D_x(z^{-1})} + \frac{S_2}{D_\omega(z)D_\omega(z^{-1})} + S_n \qquad \text{(II.6)}$$

where S_1 and S_2 are given by (II.4) using the appropriate autocorrelation values for $\{x_t\}$ and $\{\omega_t\}$, S_n is the spectrum of n_t and

$$D_x(z) = 1 - \sum_{i=1}^{L} a_i z^i \ , \ D_\omega(z) = 1 - \sum_{i=1}^{M} b_i z^i \ .$$

Putting (II.6) under a common denominator, it becomes obvious that $S_Y(z)$ is the spectrum of an autoregressive moving average process of orders L+M and L+M, i.e., AR(L+M)/MA(L+M). This is a process with L+M zeros and L+M poles, where, from equation (II.6), the numerator polynomial coefficients are related to $\{a_i\}$, $\{b_i\}$ and S_n in an obvious manner.

It can be seen that the estimation of $S_Y(z)$ using a purely AR technique (all-pole) is equivalent to approximating the (L+M) order moving average component by an AR one. This, theoretically would require an infinite order model. Finite order models of order L, however, will estimate $S_Y(f)$ with good resolution, with L depending on the various parameters of signal noise and interference, notably their relative power.

It is obvious now that whereas the resolution of Blackman and Tukey type spectral estimations are only determined by the window bandwidths in a predictable fashion, those of the AR and generally ARMA estimators are very much data dependent, requiring larger all-pole orders or ARMA modeling. Figure (II.1) shows the spectrum of a noiseless AR signal. Figure (II.2) shows the calculated AR (from exact values of the autocorrelation function) spectrum of the same signal in the presence of white noise, using order L = 20. The degradation of spectral resolution is obvious. A different approach to the demonstration of the dependence of spectral resolution on the signal-to-noise ratio is to consider a parallel resonator model for the measured signal. This is discussed in the next section.

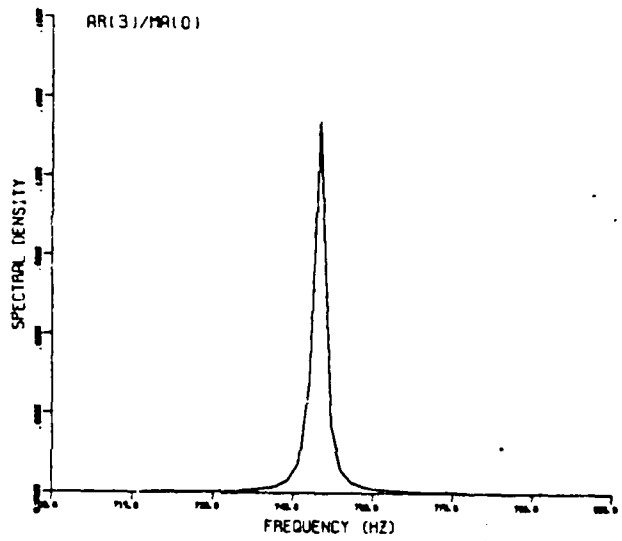
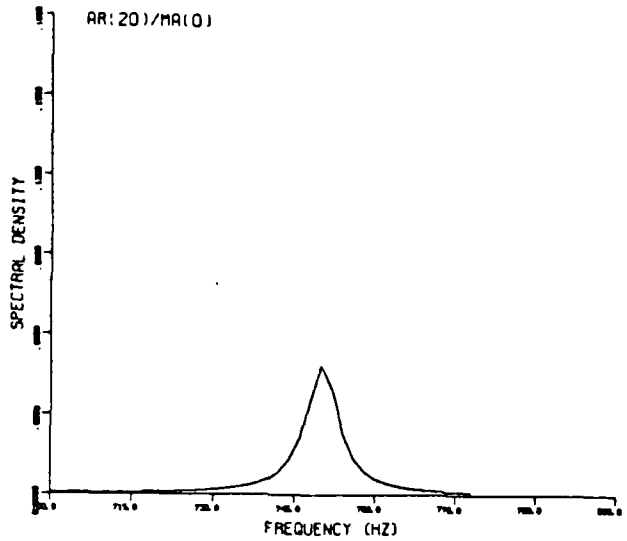
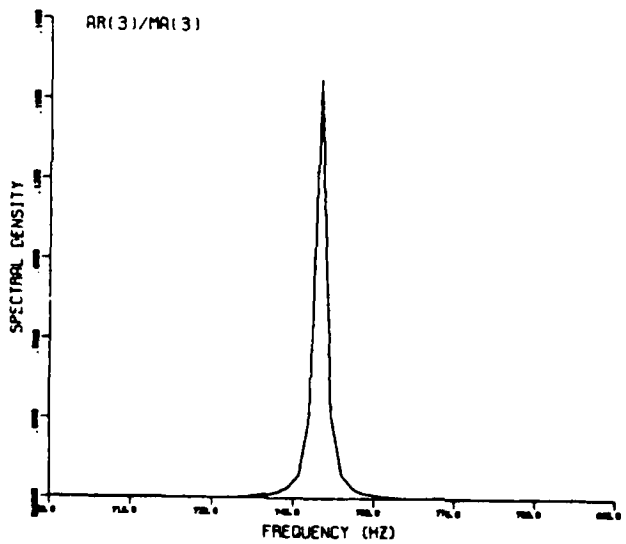


FIGURE II.1 Calculated power spectrum for $r_k = \exp(-2\pi|k|\Delta T)\cos(2\pi x 750 k\Delta T)$, $\Delta T = 1/2048$.



(a)



(b)

FIGURE II.2 Calculated power spectrum for $r_k=\delta_{ok}+\exp(-2\pi|k|\Delta T)\cdot\cos(2\pi x750k\Delta T)$.

II.3 Parallel Resonator Model

High resolution spectral estimation is normally used in situations where the signal has a "peaky" spectrum. Therefore, one may postulate the signal model as the sum of the outputs of M second-order resonators driven by white noise. This model, as will be shown in the sequel, lends itself to an investigation of the resolution degradation or bandwidth expansion as a function of measurement noise.

Mathematically, the signal s(n) is modeled as [9],

$$s(n) = \sum_{m=1}^{M} s_m(n)$$

where

$$s_m(n) = a_{m1}s_m(n-1) + a_{m2}s_m(n-2) + a_{m0}u(n) \qquad \text{(II.7)}$$

where u(n) is the driving noise of the signal generation process. It is assumed that u(n) is white with variance σ_u^2. The signal model is shown in figure (II.3).

The transfer function of the signal generation process is defined as

$$H(z) = \frac{S(z)}{U(z)} \qquad \text{(II.8)}$$

where S(z) and U(z) denote the z-transforms of the signal s(n) and the noise u(n) respectively. The transfer function $H_m(z)$ of the m'th resonator is obtained by taking the z-transform of $s_m(n)$

$$S_m(z) = a_{m1}z^{-1}S_m(z) + a_{m2}z^{-2}S_m(z) + a_{m0}U(z)$$

Hence,

$$H_m(z) = \frac{S_m(z)}{U(z)} = \frac{a_{m0}}{1 - a_{m1}z^{-1} - a_{m2}z^{-2}} \; . \qquad \text{(II.9)}$$

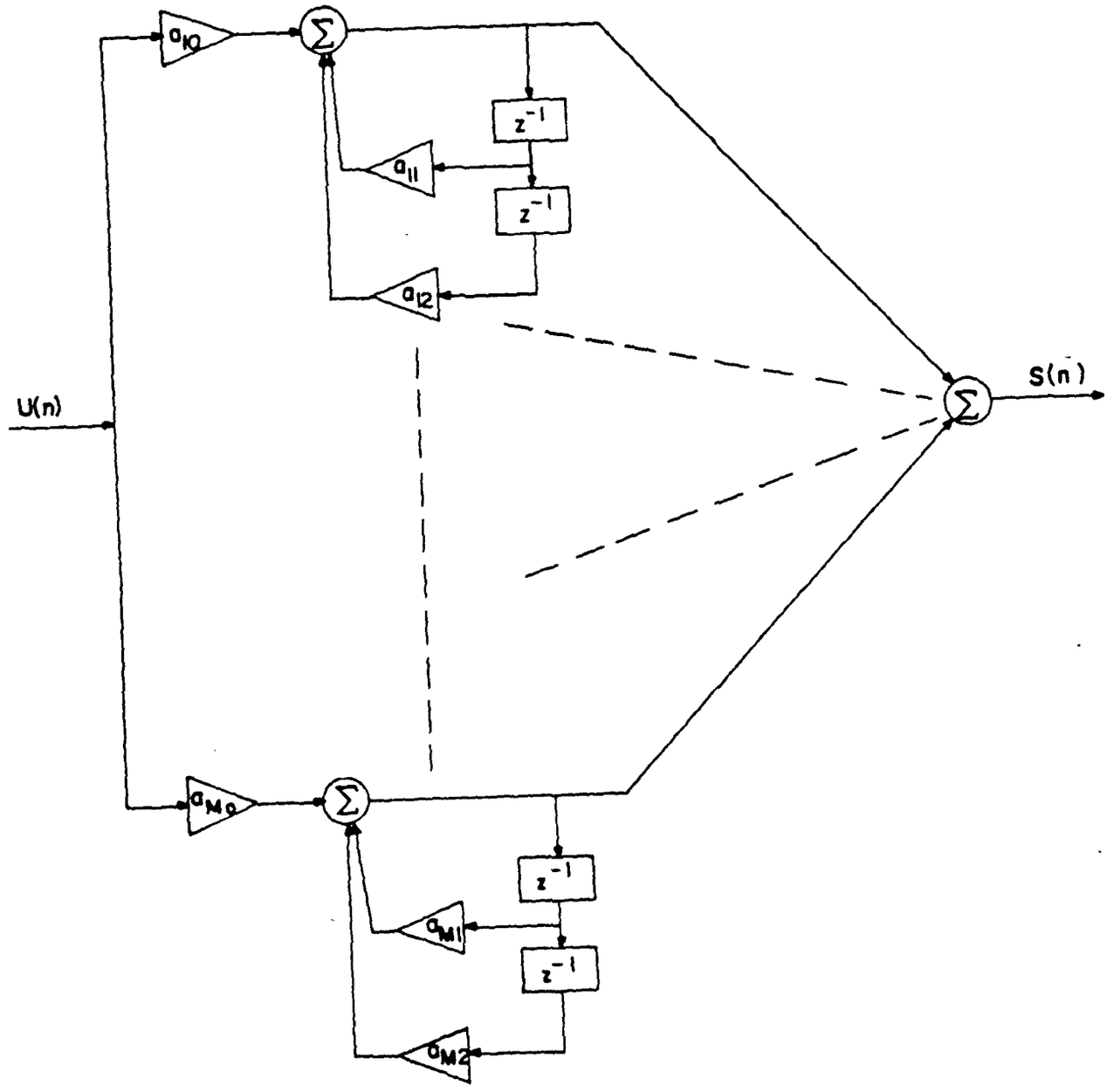


FIGURE II.3 THE SIGNAL MODEL

The z-transform of the signal s(n) is then

$$S(z) = \sum_{m=1}^{M} S_m(z)$$

$$= \sum_{m=1}^{M} U(z)H_m(z)$$

We can, therefore, express the transfer function for the generation of s(n) as:

$$H(z) = \frac{\sum_{k=1}^{M} a_{k0} \prod_{\substack{m=1 \\ m\neq k}}^{M} (1 - a_{m1}z^{-1} - a_{m2}z^{-2})}{\prod_{m=1}^{M} (1 - a_{m1}z^{-1} - a_{m2}z^{-2})} \qquad \text{(II.10)}$$

It can easily be seen that the order of the numerator polynomial N(z) of the transfer function H(z) is 2M-2 while the order of the denominator polynomial is 2M. H(z) is also recognized as the transfer function of an autoregressive-moving average (ARMA) model. Therefore, using the parallel resonator model is the same as using an ARMA model where the difference in order between the autoregressive (AR) and moving average (MA) terms is two.

Thus, in the absence of noise, s(n) can accurately be modeled as an ARMA process (or the output of a pole-zero filter). If s(n) is corrupted by the uncorrelated white noise sequence w(n), the z-power spectrum of the observation, y(n), is given by

$$S_y(z) = \left|\frac{N(z)}{D(z)}\right|^2 \sigma_u^2 + \sigma_w^2$$

$$= \frac{\sigma_u^2 |N(z)|^2 + \sigma_w^2 |D(z)|^2}{|D(z)|^2}$$

therefore, in the presence of noise, the model of y(n) is equivalent to an ARMA (2M, 2M) process. This of course is

consistent with the model derived in the preceding section directly from the addition of AR processes and white noise.

II.4 Analysis of a Single Resonator

Some insight into the behavior of the parallel resonator model in the presence of noise can be gained by examining the behavior of each resonator individually.

To simplify the analysis, we assume that the noise w(n) is added equally to the output of each resonator. In the following first the set of conditions which the parameters of each resonator must satisfy for stability and resonance are discussed. The relations between the parameters and the location of the poles of a resonator are then derived. The effect of additive noise on the poles is then examined. Finally, the perturbations of the pole positions of the noisy resonator as a function of the signal-to-noise ratio is analyzed.

The transfer function of the m'th resonator is given by

$$H_m(z) = \frac{S_m(z)}{U(z)}$$

$$= \frac{a_{m0}}{1 - a_{m1}z^{-1} - a_{m2}z^{-2}} \qquad \text{(II.11)}$$

$$= \frac{a_{m0}}{A_m(z)}$$

The poles of $H_m(z)$ must be complex conjugates for $S_m(n)$ to be a narrow band process. The locations of the poles are the solution of the quadratic equation

$$z^2 A_m(z) = z^2 - a_{m1}z - a_{m2} = 0 \qquad \text{(II.12)}$$

or

$$z = \frac{a_{m1} \pm \sqrt{a_{m1}^2 + 4a_{m2}}}{2}$$

the condition for having complex conjugate poles is that the discriminant must always be negative. That is,

$$a_{m1}^2 + 4a_{m2} < 0$$

or

$$a_{m2} < -\frac{a_{m1}^2}{4} \tag{II.13}$$

If equation (II.13) is satisfied, then the poles would be in polar coordinates,

$$z_m = \rho_m e^{j\omega_m}$$

where ρ_m is the distance from the origin of the unit circle to location of the pole and ω_m is the frequency of resonance in radians/second.

The system is stable if z_m and z_m^* are located within the unit circle. The conditions for stability are given by the Jury test

$$A_m(1) > 0$$

$$A_m(-1) > 0$$

and $\quad 1 > |a_{m2}|$

Therefore, the parameters must satisfy

$$a_{m1} + a_{m2} < 1$$

$$a_{m1} - a_{m2} > -1 \tag{II.14}$$

and $\quad |a_{m2}| < 1$

When equations (II.13) and (II.14) are satisfied, the impulse response $h_m(n)$ is easily shown to be [10]

$$h_m(n) = Z^{-1}\{H_m(z)\}$$

$$= a_{m0} \frac{\rho_m^n \sin(n+1)\omega_m}{\sin \omega_m}$$

Upon substituting for z_m and z_m^* in equation (II.12) and equating $A_m(z_m)$ and $A_m(z_m^*)$, it follows easily that

$$a_{m1} = 2r_m \cos\omega_m$$

$$a_{m2} = -r_m^2 \qquad \text{(II.15)}$$

It has already been shown that, in the absence of noise, the m'th resonator has 2 complex conjugate poles located within the unit circle and 2 zeros located at the origin of the unit circle (equation II.11). The effect of the addition of the white noise is to move the zeros from the origin towards the poles. When the noise dominates the signal, the zeros cancel the poles resulting in a flat spectrum.

If we let $y_m(n)$ be the output of the m'th noisy resonator, then

$$y_m(n) = s_m(n) + \frac{w(n)}{M}$$

$$= s_m(n) + w'(n) \qquad \text{(II.16)}$$

where $w'(n)$ is white with variance

$$\sigma_{w'}^2 = \frac{\sigma_w^2}{M^2}$$

Then, the power spectrum of $y_m(n)$ is

$$S_{y_m}(z) = \sigma_u^2 \left| H_m(z) \right|^2 + \sigma_{w'}^2$$

$$= \frac{a_{m0}^2 \sigma_u^2}{A_m(z) A_m(z^{-1})} + \sigma_{w'}^2$$

$$= \frac{a_{m0}^2 \sigma_u^2 + \sigma_{w'}^2 A_m(z) A_m(z^{-1})}{A_m(z) A_m(z^{-1})}$$

If we let

$$\sigma_\eta^2 B_m(z) B_m(z^{-1}) = a_{m0}^2 \sigma_u^2 + \sigma_{w'}^2 A_m(z) A_m(z^{-1}) \qquad \text{(II.17)}$$

we see that $y_m(n)$ may be modeled as the output of a rational transfer function with $B_m(z)$ as the numerator and $A_m(z)$ as the denominator polynomials with $\eta(n)$ as a white input. Then $B_m(z) = 1 - b_{m1}z^{-1} - b_{m2}z^{-2}$ and $\eta(n)$ is a zero mean white sequence with variance σ_η^2. It is known [11] that a $B_m(z)$ exists which has its roots on or within the unit circle. Equation (II.17) can be written as

$$\begin{aligned} &\sigma_\eta^2 (1 - b_{m1}z^{-1} - b_{m2}z^{-2})(1 - b_{m1}z - b_{m2}z^2) = \\ &a_{m0}^2 \sigma_u^2 + \sigma_{w'}^2 (1 - a_{m1}z^{-1} - a_{m2}z^{-2})(1 - a_{m1}z - a_{m2}z^2) \end{aligned} \qquad \text{(II.18)}$$

Upon expanding the above equation and equating terms of equal degree on both sides we get

$$a_{m0}^2 \sigma_u^2 + \sigma_{w'}^2 (1 + a_{m1}^2 + a_{m2}^2) = \sigma_\eta^2 (1 + b_{m1}^2 + b_{m2}^2) \qquad \text{(a)}$$

$$a_{m1} \sigma_{w'}^2 (a_{m2} - 1) = b_{m1} \sigma_\eta^2 (b_{m2} - 1) \qquad \text{(II.19)} \quad \text{(b)}$$

$$a_{m2} \sigma_{w'}^2 = \sigma_\eta^2 b_{m2} \qquad \text{(c)}$$

Equations (II.19) (b) and (c) imply

$$b_{m1} = \frac{a_{m1}(a_{m2} - 1)b_{m2}}{a_{m2}(b_{m2} - 1)} \tag{II.20}$$

Equations (II.19) (a) and (c) imply

$$\frac{a_{m0}^2\sigma_u^2 + \sigma_{w'}^2(1 + a_{m1}^2 + a_{m2}^2)}{1 + b_{m1}^2 + b_{m2}^2} = \frac{a_{m2}\sigma_{w'}^2}{b_{m2}} \tag{II.21}$$

or,

$$\frac{\lambda + (1 + a_{m2}^2 + a_{m2}^2)}{1 + b_{m1}^2 + b_{m2}^2} = \frac{a_{m2}}{b_{m2}} \tag{II.22}$$

where

$$\lambda = \frac{a_{m0}^2\sigma_u^2}{\sigma_{w'}^2}$$

If the noise variance $\sigma_{w'}^2$ is small, then $\lambda >> 1$. Equation (II.22) implies therefore that $\frac{a_{m2}}{b_{m2}} >> 1$ or $a_{m2} >> b_{m2}$ in which case equation (II.20) shows that $b_{m1} \to 0$. Then, for high signal-to-noise ratios, the zeros are located near the origin and their presence does not affect the spectrum seriously. The exact locations of the zeros is found by solving equations (II.20) and (II.19) simultaneously.

II.5 Perturbations of Pole Positions

The addition of white noise to a process given by a resonator model tends to move the power spectral density peak and to broaden the bandwidth. It has been seen in the previous section that this is caused by the zeros being shifted from the origin towards the poles. If we now model the observation again by an all-pole second-order resonator model, the positions of the poles will have changed.

It is this perturbation which causes an expansion in the estimated bandwidth. In the following this effect is examined.

The output of the m'th resonator is

$$y_m(n) = s_m(n) + w'(n)$$

where $s_m(n)$ is a second-order process,

$$s_m(n) = a_{m1}s_m(n-1) + a_{m2}s_m(n-2) + a_{m0}u(n)$$

$y_m(n)$ may also be modeled as a second-order process,

$$y_m(n) = a'_{m1}y_m(n-1) + a'_{m2}y_m(n-2) + a'_{m0}u(n).$$

In the absence of noise, a'_{m1}, a'_{m2} and a'_{m0} are identical to a_{m1}, a_{m2} and a_{m0} respectively. The autocorrelation function $r_{y_m}(n)$ of the process $y_m(n)$ has been shown to be given by [10],

$$r_{y_m}(n) = a_{m0}^2\sigma_u^2\,\rho_m^{|n|}\alpha\cos(|n|\omega_m - \phi_m) + \sigma_{w'}^2\delta(n)$$

where,

$$\alpha = \frac{1+\rho_m^2}{1-\rho_m^2}\,\frac{1+\left(\frac{1-\rho_m^2}{1+\rho_m^2}\right)^2\cot^2\omega_m}{1-2\rho_m^2\cos 2\omega_m+\rho_m^4} \tag{II.23}$$

$$\phi_m = \arctan\frac{1-\rho_m^2}{1+r_m^2}\cot\omega_m. \tag{II.24}$$

and, ρ_m and ω_m are related to the parameters a_{m1}, a_{m2} by

$$\rho_m = \sqrt{-a_{m2}}$$

$$\omega_m = \arccos\frac{a_{m1}}{2\sqrt{-a_{m2}}}$$

a'_{m1}, a'_{m2} satisfy the Yule-Walker equations resulting in

$$a'_{m2} = \frac{r_{y_m}(0)r_{y_m}(2) - r^2_{y_m}(1)}{r^2_{y_m}(0) - r^2_{y_m}(1)} = -\rho'^2_m . \tag{II.25}$$

Define the signal-to-noise ratio ξ as the power in the process $s_m(n)$ divided by the noise power, or

$$\xi = \frac{r_{y_m}(0) - \sigma^2_{w'}}{\sigma^2_{w'}}$$

$$\approx \frac{r_{y_m}(0)}{\sigma^2_{w'}} - 1 \tag{II.26}$$

$$= \gamma - 1$$

Substitute for $r_{y_m}(0)$, $r_{y_m}(1)$ and $r_{y_m}(2)$ in equation (II.25). It can be shown that (Appendix A)

$$\frac{\rho'_m}{\rho_m} = \frac{c_2\xi^2 + c_1\xi + c_0}{d_2\xi^2 + d_1\xi + d_0} \tag{II.27}$$

where,

$$c_2 = \cos^2(\omega_m - \phi_m) - \cos\phi_m\cos(2\omega_m - \phi_m)$$

$$c_1 = 2\cos^2(\omega_m - \phi_m) - 3\cos\phi_m\cos(2\omega_m - \phi_m)$$

$$c_0 = \cos^2(\omega_m - \phi_m - 2\cos\phi_m\cos(2\omega_m - \phi_m)$$

$$d_2 = \cos^2(\omega_m - \phi_m) - \cos\phi_m\cos(2\omega_m - \phi_m)$$

$$d_1 = 2(\cos^2\phi_m + \cos^2(\omega_m - \phi_m) - \cos\phi_m\cos(2\omega_m - \phi_m))$$

$$d_0 = 3\cos^2\phi_m - \cos\phi_m\cos(2\omega_m - \phi_m) + \cos^2(\omega_m - \phi_m)$$

The distance ρ_m from the origin of the unit circle to the location of the pole is related to the 3 dB bandwidth B_m in Hz of the m'th resonator by [9]

$$\rho_m = e^{-\frac{B_m}{2}}$$

then,

$$\frac{\rho_m'}{\rho_m} = \frac{e^{-\frac{B_m'}{2}}}{e^{-\frac{B_m}{2}}}$$

where B_m' is the 3 dB bandwidth in Hz of the noisy resonator. Define the bandwidth expansion factor (BEF) as

$$\text{BEF} = \frac{B_m'}{B_m} = 1 + \frac{\ln\left(\frac{\rho_m'}{\rho_m}\right)}{\ln \rho_m} \qquad \text{(II.28)}$$

Equation (II.27) shows that the BEF depends on the signal-to-noise ratio (SNR), ξ, ρ_m and ω_m. Families of curves of the BEF as a function of the SNR in the range -30 dB to +30 dB have been generated for 3 different frequencies, .125 Hz, .166 HZ and .25 Hz. They are shown in figures (II.4), (II.5) and (II.6) respectively. It is seen from the graphs that, for the same SNR, the BEF is an increasing function of ρ_m and a decreasing function of ω_m. For small ρ_m's, the resonator has a large bandwidth and adding white noise (white noise has flat spectrum between -.5 Hz and +.5 Hz) does not affect the spectrum as much as for large ρ_m's.

The above discussion demonstrates clearly that the asymptotic resolution of an all-pole model depends in a non-linear fashion on both the bandwidth and frequency of the noiseless signal.

Therefore, in general an increase in the model order does not uniformly improve the spectral estimates of all the spectral peaks and the improvement is rather unpredictable. Curves such as given in figures (II.4) - (II.6) may, however, be used as a general guideline for the amount of degradation in a given application.

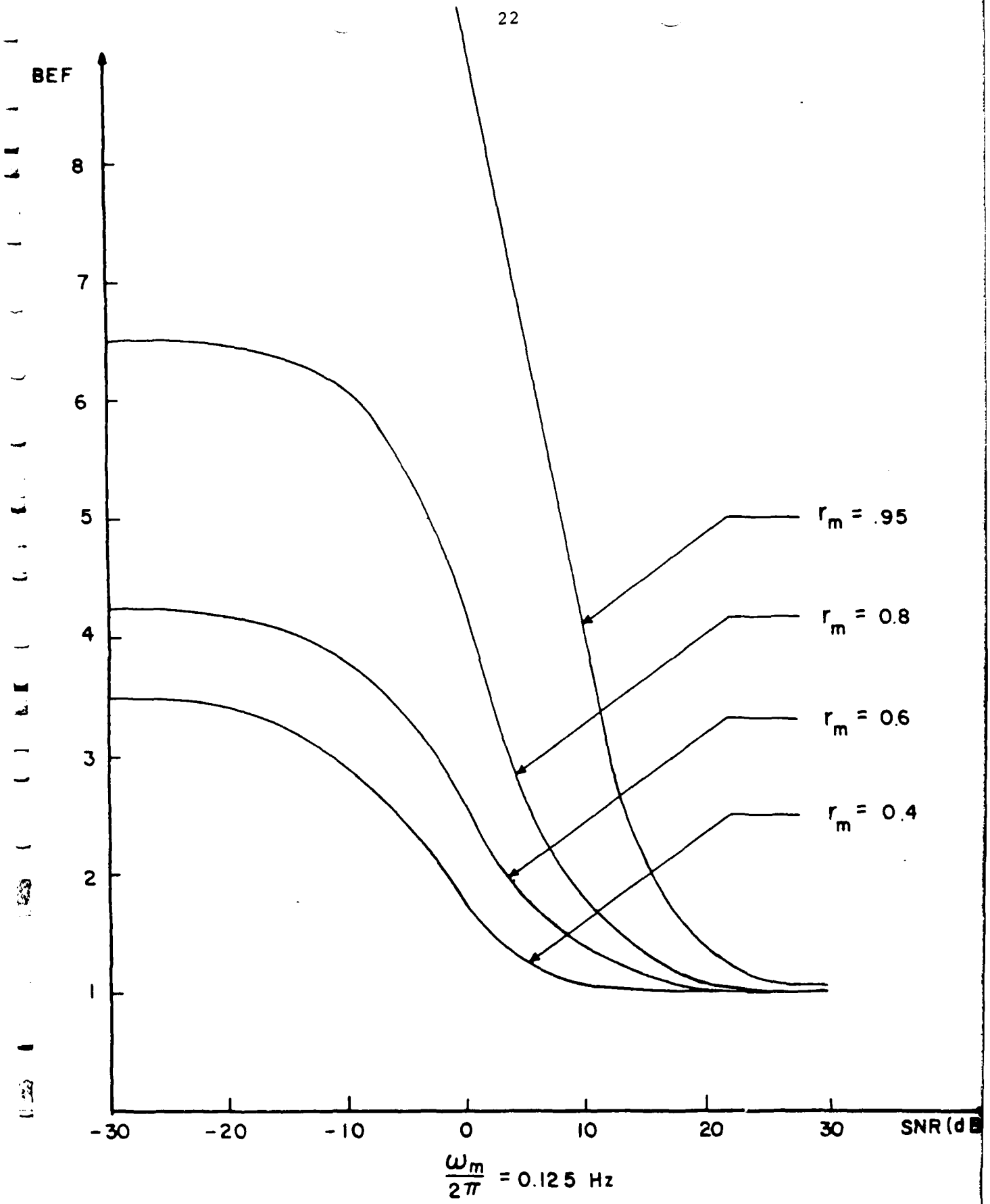


FIGURE II.4 BANDWIDTH EXPANSION FACTOR VERSUS SIGNAL-TO-NOISE RATIO FOR $f_m = .125$

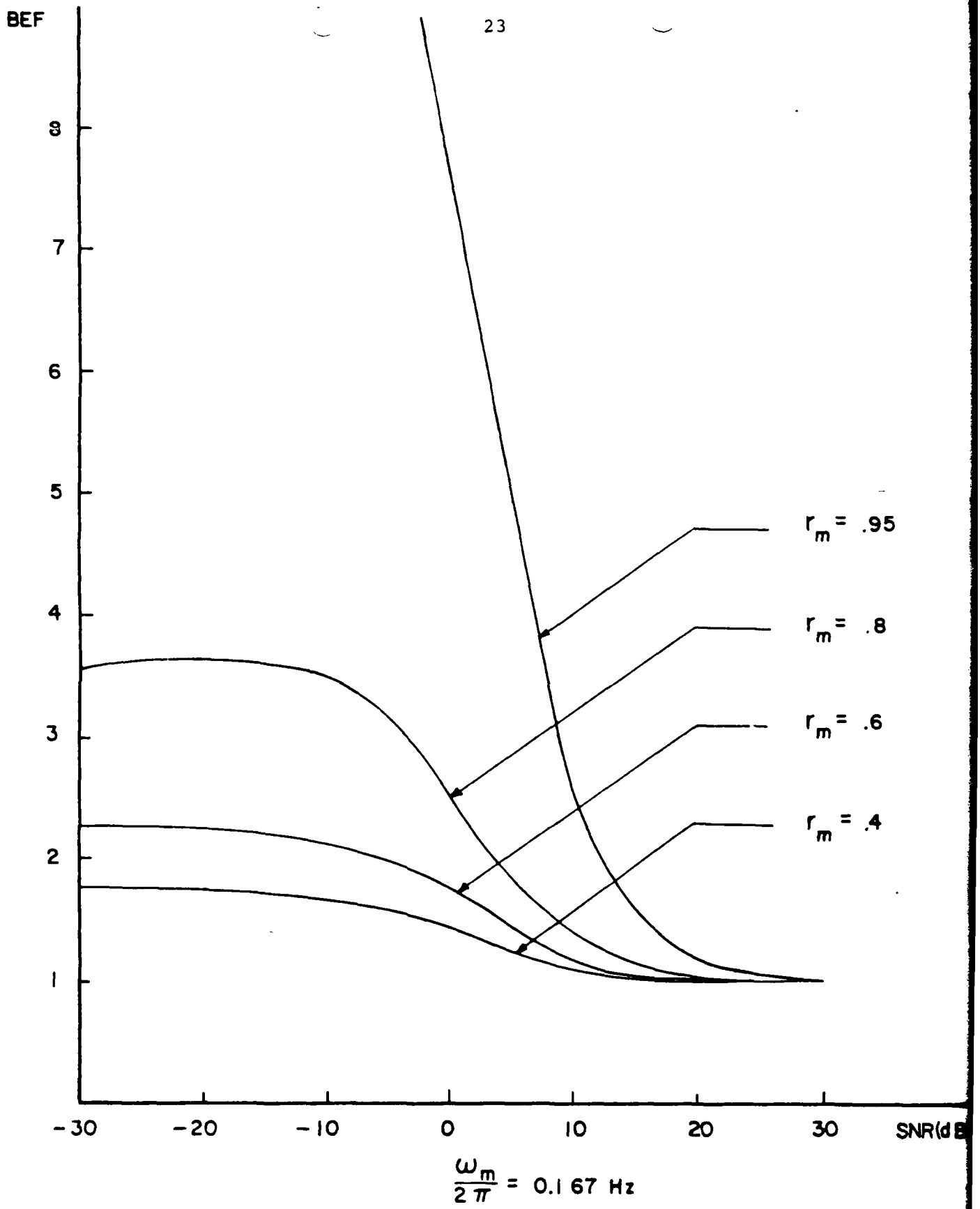


FIGURE II.5 BANDWIDTH EXPANSION FACTOR VERSUS SIGNAL-TO-NOISE RATIO FOR $f_m = .167$

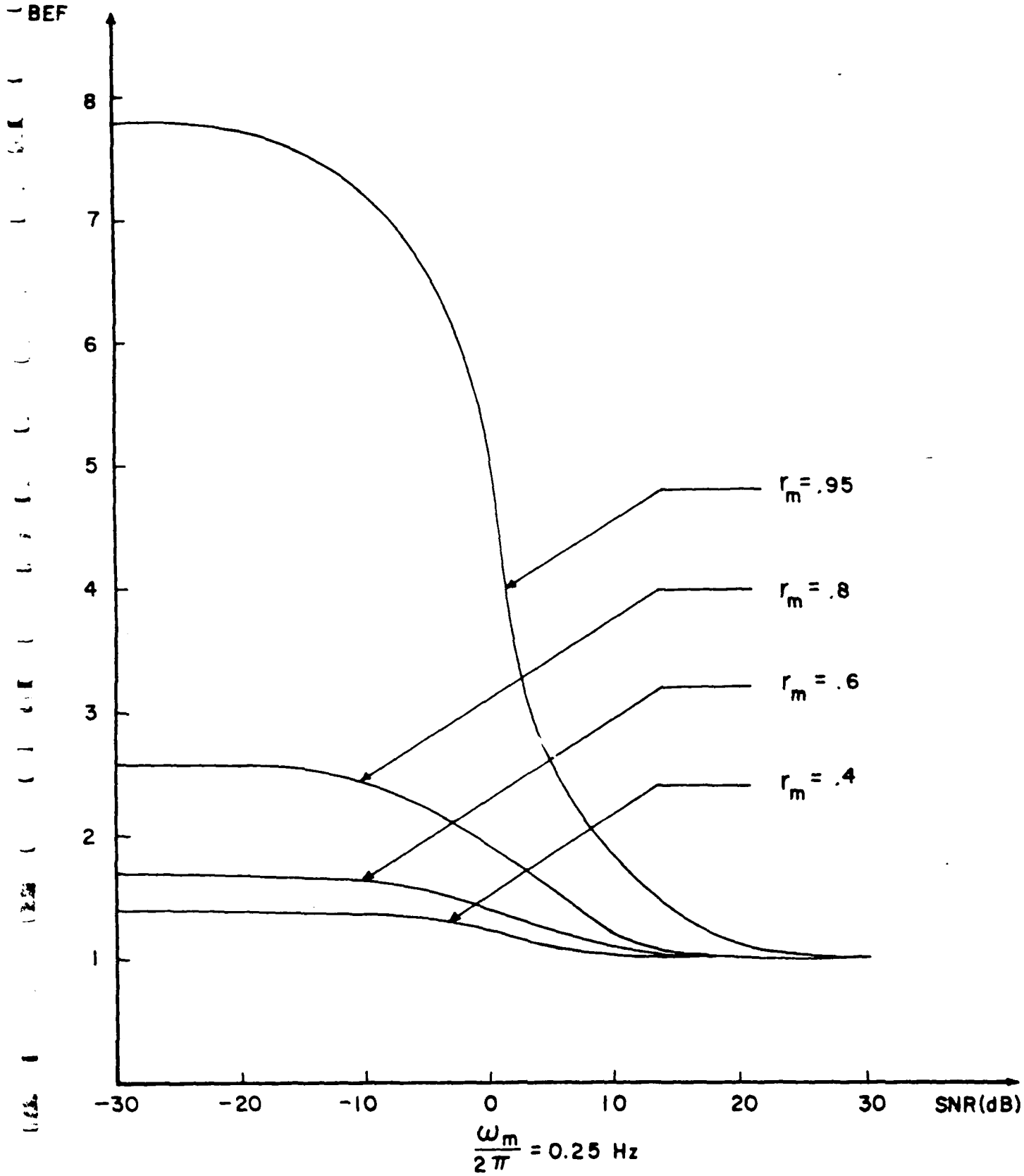


FIGURE II.6 BANDWIDTH EXPANSION FACTOR VERSUS SIGNAL-TO-NOISE RATIO FOR f_m=.25H

III. ORDER DETERMINATION FOR AR MODELS

The previous chapter demonstrated the need for increased order of an AR spectral estimator to improve resolution in the presence of noise, resulting in a larger variance for the estimates. In fact, the trade-off between the bias (or resolution) and variance of a spectral estimator is the central issue in spectral estimation by any method. For the traditional (Blackman and Tukey type) spectral estimators, this trade-off is reflected in the choice of the spectral window type and the maximum lag of autocorrelation function used. This subject, referred to as window carpentering, is discussed in detail by Jenkins and Watts [12], and is straightforward because resolution is well-defined in terms of the spectral window bandwidth.

With the popularity of data adaptive (notably the autoregressive (AR)) spectral estimation methods, similar resolution-variance trade-offs are in order. Specifically, well-defined methods are needed to determine the order of the (AR) spectral estimator for a given data sample. Furthermore for practical applications, these methods need to be on-line and as much as possible objective in nature. This problem is complicated, however, due to the data dependent nature of the resolution of the AR spectral estimator as shown before (e.g., no well-defined window bandwidth). Therefore, the question of order determination for the spectral estimator seems to be best posed as a procedure for obtaining a compromise between the AR model fit and the variance of the estimated AR parameters as a function of the model order.

Akaike [13,14] and Parzen [15] have recently introduced some methods for automatic determination of orders of autoregressive

processes. One method, based on Akaike's Information Criterion (AIC), has gained special popularity. In this report, we follow the derivations on which AIC is based, introduce appropriate modifications to account for practical estimation procedures and derive a new information criterion designated the Conditional AIC (CAIC). We then present the results of a number of numerical simulations that compare the performances of AIC and CAIC for moder order identification and spectral estimation.

III.1 Akaike's Information Criterion

Akaike derived his information criterion, AIC, as an estimate of the asymptotic relative goodness of fit of the model to the observation. Although his derivations were based on information theoretic arguments, the resulting parameters were the same as the maximum likelihood estimates. In this section we review the steps involved in obtaining AIC as they pertain to the derivation of the new criterion. We assume the time series to be described by

$$x_t = \sum_{i=1}^{L} a_i x_{t-i} + u_t, \quad t=0,\ldots N \tag{III.1}$$

$$x_{-L},\ldots,x_0 = 0$$

where u_t is zero-mean white and Gaussian and L is to be determined. Through asymptotic arguments, Akaike defines an information criterion, related to the maximum likelihood of the estimates of a_i, $\hat{a}_i$, as:

$$\mathrm{AIC}(\hat{A}) = (-2)\ln \text{ (maximum likelihood)}$$

$$+ E_\infty N \cdot ||\hat{\underline{A}} - \underline{A}||^2 \tag{III.2}$$

where E_∞ denotes asymptotic expectation, $\hat{\underline{A}}$ and $\underline{A}$ are L x 1 vectors of the coefficients a_i and their estimates $\hat{a}_i$. The practical AIC which is related to the full-information likelihood function of a Gaussian process is then given by

$$\text{AIC}(L) = N \ln (\text{MLE of innovation variance}) + 2L \qquad \text{(III.3)}$$

and the order L is chosen that minimizes AIC(L).

III.2 The New Criterion

Since the exact maximum likelihood (full information maximum-likelihood) estimates are generally not available, the conditional MLE, one based on Yule-Walker equations or Burg's algorithm, of the innovation variance are normally used in (III.3). We propose using the conditional maximum likelihood (CML) function in (III.2). This function is based exactly on the available data and we believe is a more sensitive indicator of the behavior of the estimates used in practice. Thus, in the following, the CML estimate of $\underline{A}$ and its covariance function are considered, in order to obtain tractable expressions for (III.2).

The conditional (partial information) likelihood function for the time series in (III.1) is given by:

$$L(\underline{A},\sigma_u^2,x_1,\ldots,x_L) = \frac{1}{(2\pi\sigma_u^2)^{\frac{N-L}{2}}} \exp\left(\frac{-\underline{C}^T D \underline{C}}{2\sigma_u^2}\right) \qquad \text{(III.4)}$$

where σ_u^2 is the variance of the innovation sequence u_t,

$$\underline{C}^T = [1,-a_1,-a_2,\ldots,-a_L] \text{ and}$$
$$D_{ij} = D_{ji} = \sum_{k=L+1}^{N} x_{k-i+1}\, x_{k-j+1} \,. \qquad \text{(III.5)}$$

Furthermore, the CML estimate of σ_u^2 is given by:

$$\hat{\sigma}_u^2 = \hat{\underline{C}}^T \underline{D}\hat{\underline{C}}/(N-L) \tag{III.6}$$

and a lower bound for the variance of the estimates of a_i follows from the Fisher's information matrix to be

$$\mathrm{var}[\hat{a}_i] \geq \frac{1}{N-L}\,\sigma_u^2 \Lambda^{ii} \tag{III.7}$$

where Λ^{ii} is the diagonal element of the inverse of the (L + 1)-sample covariance matrix of x_t. It is now shown that

$$\sigma_u^2 \underline{\Lambda}^{ii} = 1 \tag{III.8}$$

The innovations variance, σ_u^2, is given by the Yule-Walker relation as

$$\sigma_u^2 = r_o = [r_1 \ldots r_L]\underline{\Delta}_L^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_L \end{pmatrix} \tag{III.9}$$

where $\underline{\Delta}_{L+1}$ is the L+1-st order covariance matrix given by:

$$\underline{\Delta}_{L+1} = \begin{pmatrix} r_o & \cdots & r_L \\ r_1 & \ddots & r_{L-1} \\ \vdots & \ddots & \\ r_L & \cdots & r_o \end{pmatrix}$$

The first diagonal element of the inverse of $\underline{\Delta}_{L+1}$, Λ^{11} is then given by:

$$\Lambda^{11} = \frac{|\underline{\Delta}_L|}{|\underline{\Delta}_{L+1}|} \tag{III.10}$$

Formula (III.9) can be rewritten as

$$\sigma_u^2 = \frac{r_o|\underline{\Lambda}_L| - [r_1 \ldots r_L]\underline{S}_L \begin{bmatrix} r_1 \\ \vdots \\ r_2 \end{bmatrix}}{|\underline{\Lambda}_L|} \tag{III.11}$$

where $\underline{S}_L$ is the adjoint of $\underline{\Lambda}_L$, given by

$$\underline{S}_L = \begin{bmatrix} \mu_o & -\mu_1 & \ldots & (-1)^{L-1}\mu_{L-1} \\ -\mu_1 & & & \\ (-1)^{L-1}\mu_{L-1} & & & \mu_o \end{bmatrix}$$

and μ_i = Minor of element r_i in $\underline{\Lambda}_L$.

Comparing (III.10) and (III.11) it is obvious that we need to show

$$|\Lambda_{L+1}| = \text{Numerator in (III.11)}$$

From the form of $\underline{\Lambda}_{L+1}$ we have

$$|\underline{\Lambda}_{L+1}| = r_o|\underline{\Lambda}_L| + (-1)r_1 \sum_{i=1}^{L} r_i(-1)^{i-1}\mu_{|i-1|}$$

$$+ (-1)^2 \sum_{i=1}^{L} r_i(-1)^{i-1}\mu_{|i-2|} + \ldots$$

or

$$|\underline{\Lambda}_{L+1}| = r_o|\Lambda_L| + \sum_{j=1}^{L} \sum_{i=1}^{L} r_i r_j (-1)^{i+j-1}\mu_{|i-j|} \tag{III.12}$$

But the numerator in (III.11) is also after multiplying the matrices:

$$\text{Num} = r_o|\underline{\Lambda}_L| - \sum_{j=1}^{L} \sum_{i=1}^{L} r_i r_j \mu_{|i-j|}(-1)^{i+j-1}$$

Resulting in

$$\sigma_u^{\,2} = \frac{1}{\Lambda^{11}} \tag{III.13}$$

as required.

We now proceed to define an expression for (III.2) based on the CML estimates of $\underline{A}$ and $\sigma_u^{\,2}$. The expression for the CML given in (III.4) is now substituted in (III.2) for the maximum likelihood and using (III.8) for the second term in (III.2) and (N - L) for N we have:

$$\mathrm{CAIC}(L) = (N - L)\ln(2\pi\hat{\sigma}_u^{\,2}) + (\alpha - 1)L \tag{III.14}$$

The factor $\alpha \geq 1$ is included to account for the asymptotic nature of the criterion and the fact that (III.8) is a lower bound for the variance of $\hat{a}_1$. A similar parameter was also suggested for AIC [16] and in [14] Akaike discusses a possible approach for choosing α. Since CAIC(L) as given by (III.4) is dependent on the variance of x_t, the test is standardized by introducing a normalized innovation variance so that

$$\begin{aligned} \mathrm{CAIC}(L) &= (N - L)\ln[\hat{\sigma}_u^{\,2}/(\widehat{\mathrm{var}}\, x_t)] \\ &+ (\alpha - 1)L \end{aligned} \tag{III.15}$$

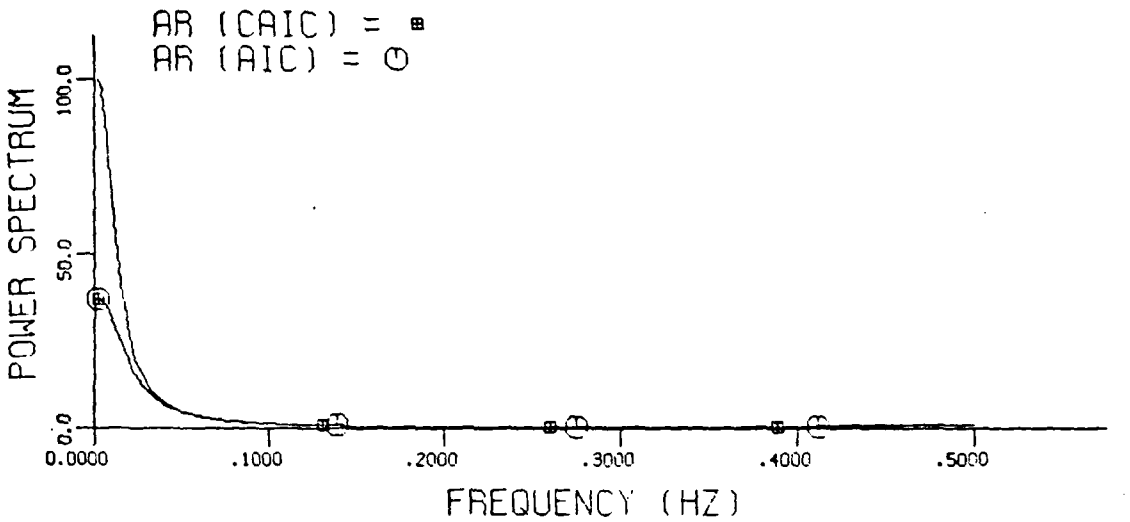
Thus CAIC(0) = 0. The factor α is chosen to give more or less weight to the error in the estimation of the parameters. In other words, resolution can be increased at the expense of the variance of the estimates by decreasing α. We have found, empirically, values of 3.5-4 to give the most stable and reasonable indication of the order.

III.3 Simulation Results

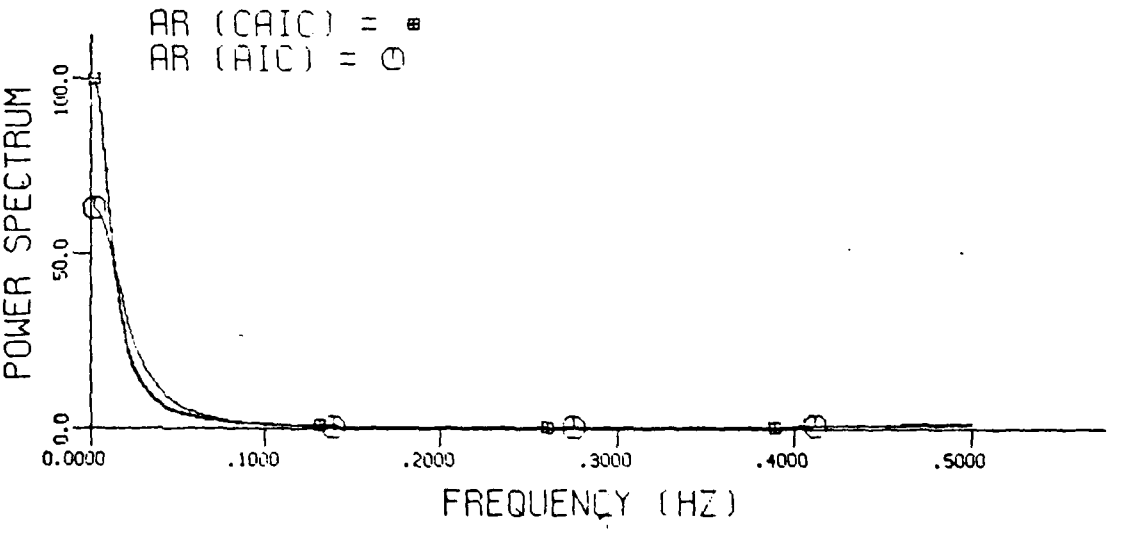
We have tested the performance of CAIC relative to AIC on a number of time series models reported previously. The data included normal as well as uniform distributions. The estimates were based on CML (least-square) and Yule-Walker methods. In the great majority of cases, CAIC performed as well or superior to AIC. Examples of these can be found in [17]. Some estimated spectra based on orders determined by AIC and CAIC are also shown in figures (III.1) - (III.3). Yule-Walker equations with auto-correlation function estimates given by

$$r_i = \frac{1}{N} \sum_{j=1}^{N-i} x_j x_{j+i}$$

were used. The example shown in Figure (III.1) indicates the relative stability of CAIC. Figure (III.2) shows that the model order chosen by AIC results in spurious peaks, while giving higher peak resolution than the CAIC based one. Figure (III.3) shows that an increase in white noise level increased the AIC order to the point that spurious spectral peaks became pronounced while CAIC remained nearly the same, showing the relative stability of CAIC.



(a)



(b)

FIGURE III.1 Power spectrum of $x_t = .4x_{t-1} + .5x_{t-2} + u_t$.

a) N=100, CAIC=2, AIC=2.

b) N=500, CAIC=2, AIC=8.

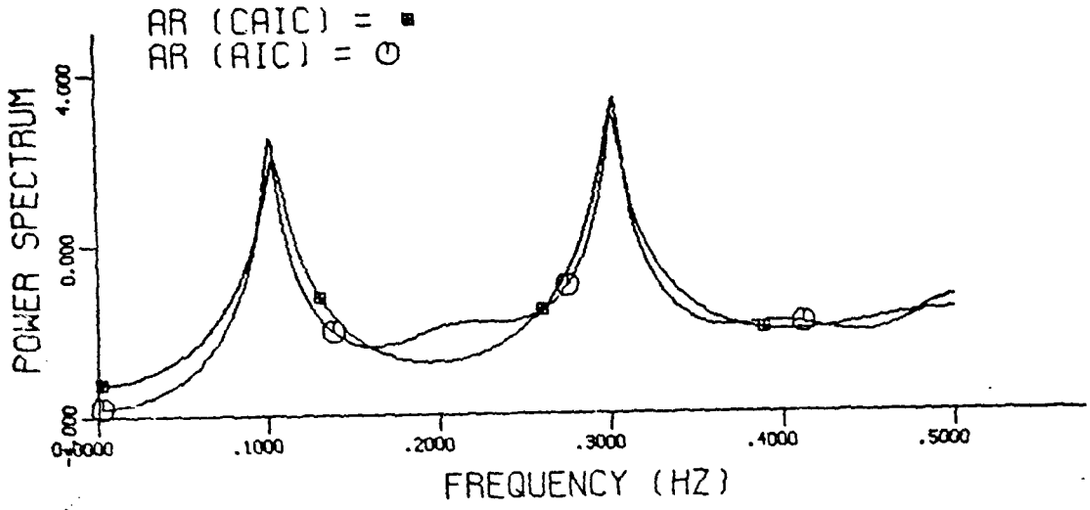


FIGURE III.2 Log power spectrum of $x_t=\sin(.6\pi t+30°)+0.707\sin(.2\pi t+60°)+n_t$, $\sigma_n^2=.15$; N=100; CAIC=6, AIC=10.

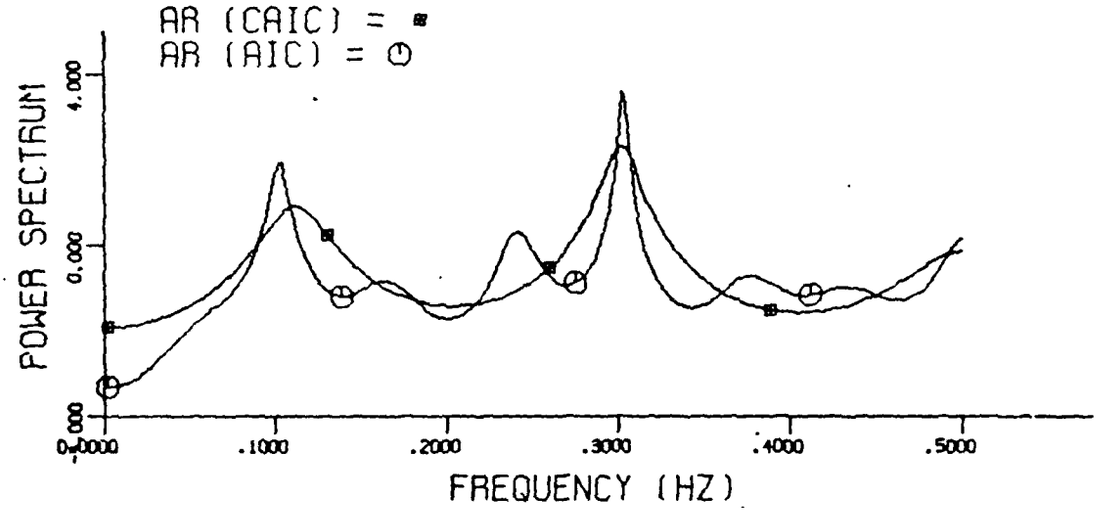


FIGURE III.3 Log power spectrum of $x_t=\sin(.6\pi t+30°)+0.707\sin(.2\pi t+60°)+n_t$; $\sigma_n^2=.316$, N=100; CAIC=5, AIC=15.

IV. SOME SUBOPTIMUM ARMA SPECTRAL ESTIMATORS

As shown in Chapter II, the presence of additive noise on the observations of an AR process implies that they are an ARMA process. A popular approach to circumvent the degradation, wrought by this model change, of the AR spectral estimate is to derive spectral estimators based on the ARMA model. Box and Jenkins [7] among others, provide an algorithm to give a close approximation to the maximum likelihood (ML) estimate of the ARMA parameters, assuming Gaussian observations. This is considered the optimum estimator in view of the desirable properties of maximum likelihood [18]. But optimum estimators have computational disadvantages, described in the next chapter, which have compelled many researchers to consider alternative criteria that yield spectral estimators with greater computational efficiency. There is presently considerable activity in this area, appropriately termed suboptimum ARMA spectral estimation.

In this chapter a new suboptimum scheme is reported for estimating the power spectral density (PSD) of an ARMA process of known orders. After a preliminary data reduction, this scheme, called the least squares (LS) estimator, minimizes a sum of squared quadratic functions of the AR coefficients using a nonlinear least squares algorithm. The poles of the estimated PSD are found from the minimizing AR coefficients, and zeros are found from quadratic functions of these coefficients. Note that these results have been published [19].

The general idea of least squares fitting the ARMA parameters is not new, and various other approaches have been suggested

(see, e.g., [7] and [20]). The scheme discussed here is, however, analogous to a minimum mean-squared error estimation of the parameters appearing in the estimator discussed in [21] and [22]. A modification of the latter estimator that is based on the modified Yule-Walker equations (the MYW estimator) is also presented, in which the problem of negative excursions of the estimated PSD is corrected by tapering the estimated moving-average autocorrelation function. Examples are shown that compare the performance of these ad hoc techniques to the approximate maximum likelihood method of Box and Jenkins.

IV.1 The Spectrum of an ARMA Process

Assume that we observe $x_t, t=1,\ldots,N$ where x_t is stationary and Gaussian of mean zero, and that x_t fits an ARMA (L,M) model. Then we can write

$$x_t - \sum_{i=1}^{L} a_i x_{t-1} = u_t \tag{IV.1}$$

where a_i are the AR parameters for the ARMA model and u_t is the MA residual sequence given by

$$u_t = \varepsilon_t - \sum_{i=1}^{M} b_i \varepsilon_{t-i} \tag{IV.2}$$

where b_i are the MA parameters and ε_t is a zero-mean uncorrelated normal sequence of variance σ_ε^2. Define

$$A^T = [1,-a_1,\ldots,-a_L].$$

Then the PSD of x_t is given by

$$S_x(z) = S_u(z)\left|A^T Z_L\right|^{-2} \tag{IV.3}$$

where $Z_k^T = [1,z^{-1},\ldots,z^{-k}]$, z being the z-transform operator

$z = e^{j\omega}$, and where $S_u(z)$ is the PSD of u_t. Writing

$$S_u(z) = \sigma_u^2[1 + \beta_1(z^{-1} + z) + \ldots + \beta_M(z^{-M} + z^M)] \tag{IV.4}$$

we can express the PSD of u_t in terms of its variance σ_u^2 and M normalized autocorrelation function coefficients β_i. The normality of u_t gives us the conditional expectation [23, pp. 218-225]

$$E(u_t|u_{t-i}) = \beta_i u_{t-i} \tag{IV.5}$$

and, in general, β_i is the best linear predictor of u_t given u_{t-i} in the minimum mean-square error (mmse) sense.

IV.2 The LS Estimator

A least squares estimate of β_i is obtained by minimizing S_i given by

$$S_i = \sum_{t=L+i+1}^{N} (u_t - \beta_i u_{t-i})^2$$

$$= \sum_{t=L+i+1}^{N} (A^T X_{[t,t-L]} - \beta_i A^T X_{[t-i,t-L-i]})^2$$

where

$$X^T_{[i,j]} = [x_i, x_{i-1}, \ldots, x_j], \; i > j.$$

The derivative with respect to β_i vanishes for

$$\hat{\beta}_i = \frac{A^T R_i A}{2A^T R_{oi} A} \tag{IV.6}$$

where

$$R_i = \sum_{t=L+i+1}^{N} [X_{[t,t-L]} X^T_{[t-i,t-L-i]}$$

$$+ (X_{[t,t-L]} X^T_{[t-i,t-L-i]})^T]$$

$$R_{oi} = \sum_{t=L+i+1}^{N} [X_{[t-i,t-L-i]} X^T_{[t-i,t-L-i]}].$$

That $\hat{\beta}_i$ minimizes S_i is seen by taking the second derivative

$$\frac{d^2}{d\beta_i^2} S_i = 2A^T R_{oi} A = 2 \sum_{t=L+i+1}^{N} (A^T X_{[t-i,t-i-L]})^2 > 0. \qquad \text{(IV.7)}$$

This also establishes a nonzero denominator for $\hat{\beta}_i$ in (IV.6).

To get an estimate of the variance of u_t we use the sample moment

$$(N-L)\hat{\sigma}_u^2 = \sum_{t=L+1}^{N} u_t^2 = \sum_{t=L+1}^{N} A^T X_{[t,t-L]} X^T_{[t,t-L]} A$$

$$= A^T R_{oo} A \qquad \text{(IV.8)}$$

where R_{oo} is given in (IV.6).

The fact that u_t is MA(M) implies that $\beta_i = 0$ for $i > M$, so one method to fit the vector A to the data is to solve the set of equations $\hat{\beta}_i = 0$, $i = M + 1, \ldots, M + L$ from (IV.6). Recognizing that the denominators are nonzero, we can instead solve the simpler system

$$A^T R_i A = 0, \quad i = M + 1, \ldots, M + L. \qquad \text{(IV.9)}$$

The locus of solutions for each of the above equations is a quadric surface [24, pp. 287-294], and there is no solution to the system if the surfaces do not jointly intersect at least one point. In order to obtain a value for A whether or not there is a common intersection, we choose A to minimize Q given by

$$Q = \sum_{i=M+1}^{M+L} (A^T R_i A)^2. \tag{IV.10}$$

The value of Q is zero at any point of intersection for all the surfaces in (IV.9). The vector A obtained via minimization of Q will henceforth be referred to as the LS estimate. Using the LS estimate, we next find $\hat{\beta}_i$, i=1, ... ,M using (IV.6) and $\hat{\sigma}_u^2$ using (IV.8) and substitute these values in (IV.4) and (IV.3) to get an estimate of the PSD of x_t, henceforth termed the LS spectral estimate.

IV.3 The MYW Spectral Estimator

The rational spectral estimator in [ ] (the MYW estimator) produces $\hat{A}$ to solve the modified Yule-Walker equations, which for an ARMA (L,M) process are

$$C_{M+1}\hat{A} = 0 \tag{IV.11}$$

where

$$C_k = \left|\left|c_{ij}\right|\right|_k = \left|\left|c_x(k+i-j)\right|\right|,$$

$$i=1, \ldots ,L; \; j=1, \ldots ,L+1,$$

and where $c_x(i)$ is an estimate of the ith lag autocorrelation of x_t, for example,

$$c_x(i) = \frac{1}{N} \sum_{t=1}^{N-i} x_t x_{t+i}. \tag{IV.12}$$

Then estimates of the autocorrelation function $r_u(i)$ of u_t are found by

$$\hat{r}_u(i) = \hat{A}^T C_i \hat{A}, \quad i=0,1, \ldots ,M \tag{IV.13}$$

and the estimated PSD is

$$\hat{S}_x(z) = [\hat{r}_u(0) + \hat{r}_u(1)(z^{-1}+z) + \ldots + \hat{r}_u(M)(z^{-M} + z^M)] \cdot \left|\hat{A}^T Z_L\right|^{-2}. \tag{IV.14}$$

Note that the LS and MYW estimates of A are asymptotically equivalent. To see this, observe that

$$p \lim_{N\to\infty} (N-L-k)^{-1} R_k = \Lambda_k + \Lambda_k^T$$

(denotes convergence in probability) [25] where Λ_k incorporates the mth lag autocorrelation $r_x(m)$ of x_t according to

$$\left|\left|\Lambda_{ij}\right|\right|_k = r_x(k+i-j),$$
$$i=1, \ldots, L+1, \; j=1, \ldots, L+1,$$

and that $p \lim_{N\to\infty} C_k = \left|\left|\Lambda_{ij}\right|\right|_k$, $i=1, \ldots, L$; $j=1, \ldots, L+1$. The solution to (IV.11) obtained by replacing C_{M+1} by its limiting value also solves (IV.9) with $(N-L-k)^{-1}R_k$ replaced by its limiting value. There is one other solution to this asymptotic form of (IV.9), but we conjecture that it is outside the stationary region for A.

The spectral estimate given by (IV.14) is not guaranteed to be a nonnegative function of $z = e^{j2\pi f}$ for $f\varepsilon[0,1/2]$. For instance, if observations consist of two additive narrow-band AR(2) signals having center frequencies in close proximity, the true z-spectrum consists of two closely spaced poles just inside the unit circle and a zero just inside the unit circle and between the poles in frequency. An error in estimating the numerator polynomial in (IV.14) can cause what should be a near-zero positive value at the bottom of the trough in the frequency response of the numerator to be a near-zero negative value, thus making the PSD estimate

negative in a region near the peak. An effective counter is to reduce the depth of the trough by multiplying the $\hat{r}_u(i)$ by a taper which slightly reduces the frequency resolution of the estimated numerator. Use of the linear taper

$$T_i = 1 - i/K, \quad i=0,1, \ldots ,M, \quad K \geq M \tag{IV.15}$$

has been successful in eliminating negative excursions of the PSD estimate when used to produce $\hat{r}_u(i)^t$ according to

$$\hat{r}_u(i)^t = T_i \hat{r}_u(i), \quad i=0, \ldots ,M.$$

Another strategy, used in [20], is to replace the numerator in (IV.14) by the periodogram of u_t. For short data records, however, the periodogram often cannot adequately represent the moving average spectrum, resulting in inaccurate indication of the power under the peaks of the ARMA spectrum. In tests of the MYW estimator, all negative excursions were eliminated for $K >> M$. No negative excursions were noted for the LS algorithm, but the existence of a guarantee has not been investigated. It is emphasized that the use of the numerator taper does not sacrifice the resolution of the ARMA spectral estimates. This is due to the fact that resolution is mainly determined by the poles and the MA spectra of interest are relatively smooth.

IV.4 Simulation Results

Figure (IV.1) shows LS estimates of the PSD of 20 realizations of $x_t = w_t + Y_t + 0.5\, n_t$ where $w_t = 0.4\, w_{t-1} - 0.93\, w_{t-2} + \varepsilon_t$ and $y_t = -0.5\, y_{t-1} - 0.93\, y_{t-2} + \eta_t$ and where ε_t, η_t, and n_t are mutually independent i.i.d. Gaussian of mean zero and unit variance. Q in (IV.10) is minimized using the conjugate gradient technique [26]

with A starting at the origin. The combined effect of the plots is to suggest a bias in the estimator which smears the spectra. In Fig. (IV.2) the tapered MYW estimate is depicted for the same set of realizations. K in (IV.15) is set to the minimum value that succeeds in eliminating all negative excursions of the spectral estimate. In this set, the untapered MYW estimates of seven realizations went negative, and K = 34. Figure (IV.3) shows the unconditional least squares Box and Jenkins estimate for the same set of realizations. It is apparent that the performance of the MYW estimator is nearly as good as that of the Box and Jenkins estimator for most realizations. The disturbing tendency of the MYW estimator to produce an occasional aberrant estimate is also exhibited in the figure.

The simulations provide evidence that the computational simplification, compared to the optimum schemes, provided by data reduction in the LS estimator is accomplished at the expense of a substantial tradeoff in statistical efficiency. The results also suggest that the MYW estimator has higher statistical efficiency than the LS estimates. These observations, coupled with the high computational efficiency of the former, lead us to conclude that the MYW estimator is the superior of the two suboptimum spectral estimators considered.

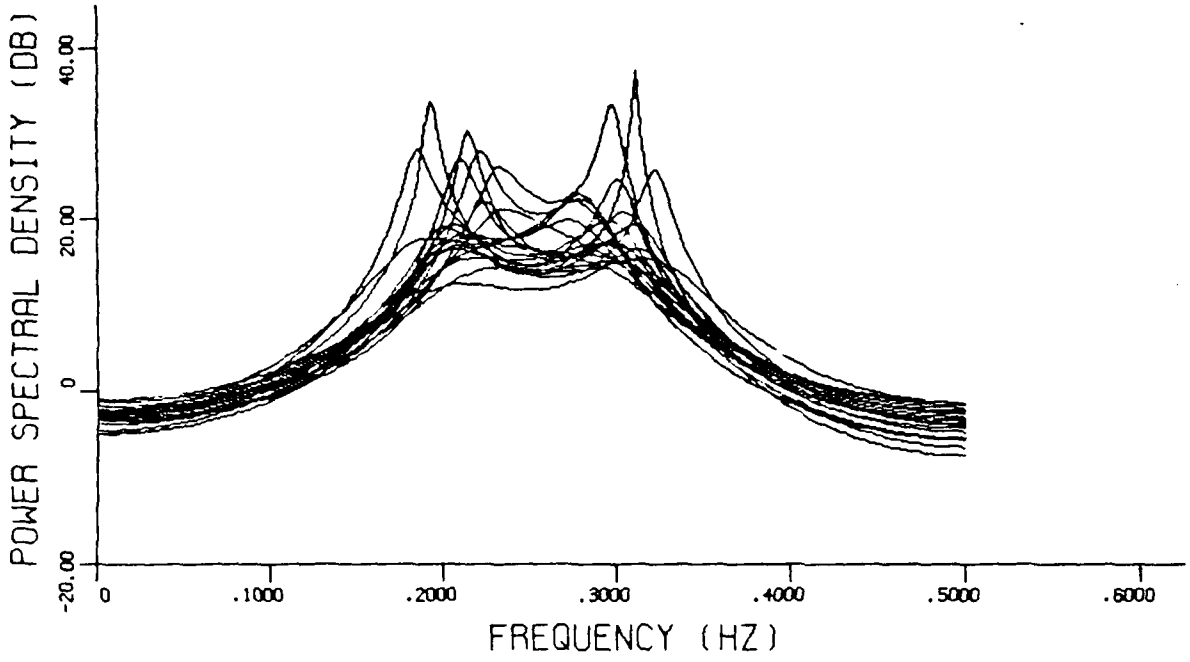


Fig. IV.1. LS spectral estimate, N = 125.

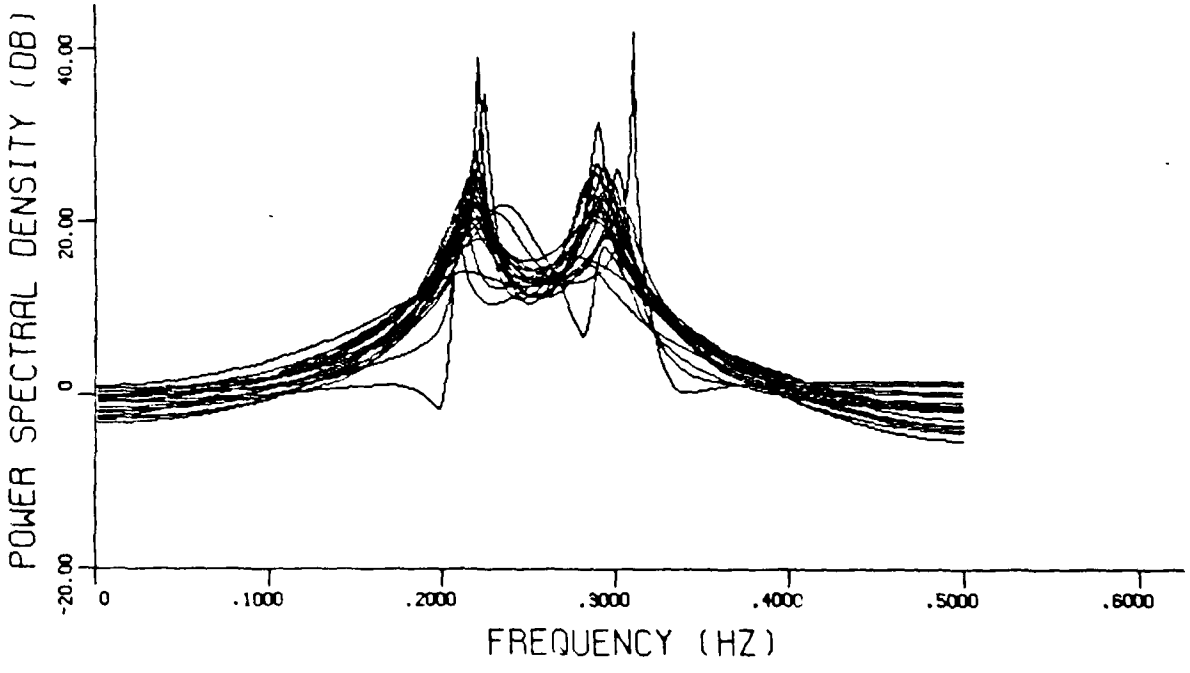


Fig. IV.2. MYW spectral estimate, N = 125.

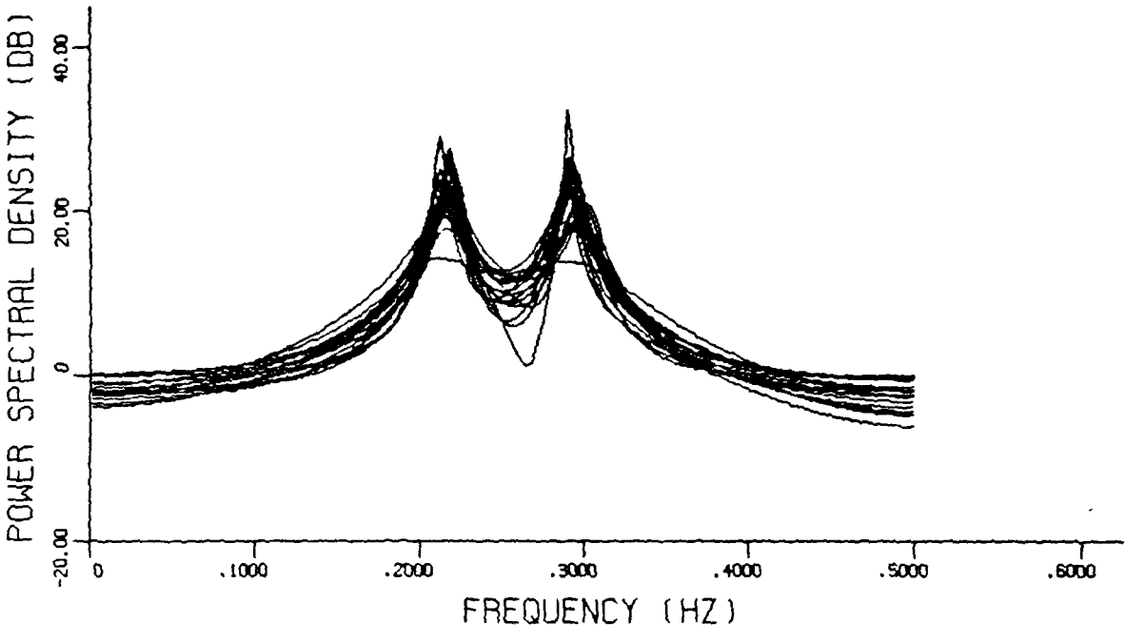


Fig. IV.3. Unconditional least squares Box and Jenkins spectral estimate, N = 125.

V. STATISTICAL CLASSIFICATION OF SOME ARMA SPECTRAL ESTIMATORS

Many of the spectral estimates appearing in the engineering and statistics literature over the last two decades fall into one of two classes. In the minimal sufficient (MS) class, the AR and MA parameters are adjusted simultaneously to give a least squares or approximately least squares fit to the observed data vector. The name "minimal sufficient" acknowledges that no reduction of the data is a sufficient statistic for stationary Gaussian time series having zeros in the power spectral density (PSD) [31]. The resulting estimates are approximately maximum likelihood (ML) for Gaussian data. In the sequel, it is assumed that the data is a Gaussian time series, and the ML ARMA parameter estimate is referred to as optimum, for reasons discussed in Chapter IV. Thus, optimum solutions are contained in the MS class. The equations for the least squares solution are highly nonlinear, and thus estimators in this class are typified by nonlinear optimization algorithms and other iterative approaches and their concomitant pitfalls, i.e., nonconvergence or convergence to local rather than global extrema, the need for preliminary ARMA parameter estimates to start the iterations, and large computer memory and time requirements. The autocorrelation function class is composed of estimators which utilize a fixed, finite number of lags of the sample ACF. The estimation equations are usually linear or quadratic, and often the AR parameters are estimated alone, after which estimates either of the MA parameters or of some function related to the PSD of the MA residuals are computed using the AR estimates. All

of the estimators in this class are suboptimum, the truncated sample ACF being a data reduction. Further, an optimum fitting requires that all parameters be adjusted simultaneously. Suboptimality, then, is the price paid for the gains in computational simplification and relaxed memory requirements of estimators in this class.

In its emphasis on computational efficiency, the literature has left largely unanswered questions regarding the statistical efficiency of estimators in the ACF class. We provide in this chapter an evaluation of Fisher's information matrix for the truncated sample ACF of ARMA processes. Only the asymptotic (infinite observation record, i.e., $N \to \infty$) case is treated to make the analysis tractable, and the magnitude and angle of poles and zeros of the PSD are taken as parameters of interest. The results will yield asymptotic bounds on the statistical efficiency of any estimator in the ACF class.

V.1 Classification of Some ARMA Spectral Estimators

In keeping with the goal of obtaining asymptotic results, the estimators to be discussed are classified according to the limiting form that the statistic on which they are based takes as N approaches infinity. The MS class includes the methods of Tretter and Steiglitz [32], Hannan [33], Akaike [34], Konvalinka and Matousek [35], and Box and Jenkins [7, pp. 231-235]. The methods of Walker [36], Hsia and Landgrebe [37], Graupe, Krause and Moore [38], Sakai and Arase [39], Satorius and Alexander [40], Kaveh [21], Kinkel et al. [22], and Bruzzone and Kaveh [19] are included in the ACF class. Cadzow's method [20] is based on a reduction of nearly all N lags of the sample ACF and hence does

not fit into either class. We have observed a tendency toward large frequency errors in Cadzow's method as well as an exaggeration of the sharpness of spectral peaks while requiring nearly as much computer time and memory as methods in the MS class. These facts, in conjunction with the suboptimality of the method as a result of the use of a data reduction as well as non-simultaneous determination of the AR and MA parameters, lead us to dismiss Cadzow's method from further analysis. A number of estimators, e.g., [36], which are based on the sample ACF yet claim to be asymptotically efficient from a subclass of the ACF class. There is no discrepancy here, however, in that efficiency obtains in the limit as first the number of observations, then the number of sample ACF lags, approach infinity. In practice we are not at liberty to extend the estimators in such fashion, so these estimators are analyzed assuming a finite truncation of the sample ACF.

Looking more closely at the ACF class, we see that [37], [21], [22], [40] and [19] first compute an estimate of the AR parameters, and then use these to estimate some function related to the MA parameters. Thus, the statistical efficiency of these estimators is determined largely by the efficiency of the AR parameter estimates in the first stage. The AR parameters determine the pole locations of the ARMA PSD, and consequently they have far greater influence on its shape than do the MA parameters in the problem of high resolution spectra. It is clear that the asymptotic efficiency with respect to only the estimated pole locations of various estimators is a meaningful criterion by which to judge them. For example, [37], [21], [22],

[40] and any method based on modifying the numerator portion of their estimated PSD's (as in Kay [41]) are equivalent by this criterion, a result that agrees with our experience.

The two popular sample ACF's

$$c_x(i) = \frac{1}{N} \sum_{t=1}^{N-i} x_t x_{t+i} \; ; \; i=0,1,\ldots,k$$

and (V.1)

$$r_x(i) = \frac{1}{N-i} \sum_{t=1}^{N-i} x_t x_{t+i} \; ; \; i=0,1,\ldots,k$$

are asymptotically equivalent for finite k. We take as the statistic of interest $C^T_{[k_1,k_2]} = [c_x(k_1), c_x(k_1+1), \ldots, c_x(k_2)]$, $k_1 < k_2$. This allows us to take into account the effect of not using low-lag ACF values, as is the case in the five estimators listed in the previous paragraph, wherein $c_x(o)$ is not used to estimate the AR parameters. The asymptotic Fisher's information matrix is derived for any subset of the poles and zeros by appropriate choice of the parameter vector.

We mention briefly the work of Gersch [42], who provides the asymptotic covariance matrix of the AR parameter estimates gotten from the modified Yule-Walker operations, and that of Sakai and Tokumaru [47], who give the covariance of the estimated power spectrum gotten from using [21] or [22] (these are equivalent). Little else has been done to analyze the suboptimum ARMA spectral estimators, and while these results are useful, they lack the general applicability of the analysis we now undertake.

V.2 Evaluation of Fisher's Information Matrix

We begin by evaluating the asymptotic covariance matrix of the sample ACF for an ARMA(L,M) process, where we assume $L > 0$, using the notation and assumptions of Chapter IV. It is known that [12, p. 181], for a stationary time series x_t,

$$\lim_{N\to\infty} \operatorname{cov}[c_x(k), c_x(\ell)] = \frac{1}{N} \sum_{i=-\infty}^{\infty} \gamma(i)\gamma(i+\ell-k)+\gamma(i+\ell)\gamma(i-k) \qquad \text{(V.2)}$$

where $\gamma(i) = E(x_t x_{t+i})$. The ACF of an ARMA(L,M) process is recursive beginning with lag M. Starting values are given by the appropriate system below (see [44]),

$$\begin{pmatrix} 1 & -a_1 & -a_2 & \cdots & -a_{L-1} & -a_L \\ -a_1 & 1-a_2 & -a_3 & \cdots & -a_L & 0 \\ -a_2 & -a_1-a_3 & 1-a_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ -a_L & -a_{L-1} & -a_{L-2} & \cdots & -a_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L-1) \\ \gamma(L) \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} D_0 \\ D_1 \\ D_2 \\ \vdots \\ D_L \end{pmatrix} ; \; L \geq M \geq 0 \qquad \text{(V.3a)}$$

$$\begin{pmatrix} 1 & -a_1 & -a_2 & \cdots & -a_{L-1} & -a_L & 0 & 0 & \cdots & 0 \\ -a_1 & 1-a_2 & -a_3 & \cdots & -a_L & 0 & 0 & 0 & \cdots & 0 \\ -a_2 & -a_1-a_3 & 1-a_4 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ -a_L & -a_{L-1} & -a_{L-2} & \cdots & -a_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -a_L & -a_{L-1} & \cdots & -a_2 & -a_1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -a_L & \cdots & -a_3 & -a_2 & -a_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L-1) \\ \gamma(L) \\ \gamma(L+1) \\ \gamma(L+2) \\ \vdots \\ \gamma(M) \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} D_0 \\ D_1 \\ D_2 \\ \vdots \\ D_L \\ D_{L+1} \\ D_{L+2} \\ \vdots \\ D_M \end{pmatrix} M>L>$$

(V.3b)

where

$$D_i = \sigma_\varepsilon^2 \sum_{j=i+1}^{M} \sum_{\ell=1}^{j-i} (a_\ell - b_\ell) B_j \det(\alpha_{j-i-\ell}) + \sigma_\varepsilon^2 B_i , \qquad (V.4)$$

with $B_o = 1$, $B_i = b_i$ for $i=1,\ldots,M$, $B_i = 0$ for $i > M$, $a_i = 0$ for $i > L$, and $\sum_i^j = 0$ for $j < i$. Also,

$$\alpha_k = \begin{cases} \begin{pmatrix} a_1 & a_2 & \cdots & a_k \\ -1 & a_1 & \cdots & a_{k-1} \\ 0 & -1 & \cdots & a_{k-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_1 \end{pmatrix} & ; k \geq 1 \\ 1 & ; k = 0 \end{cases} \qquad (V.5)$$

where $a_i = 0$ for $i > L$. For $M > 1$ and $L > 1$ we extend the vector of starting values to lag M+L-1 using the AR recursion

$$\gamma(k) = \sum_{i=1}^{L} a_i \gamma(k-i). \qquad (V.6)$$

Although this recursion can be used to extend the ACF arbitrarily far, the infinite sums in (V.2) are simplified by expressing the ACF in terms of its poles, i.e.,

$$\gamma(i) = \sum_{j=1}^{L} \lambda_j G_j^{\,i-M} , \; i \geq M , \qquad (V.7)$$

where the G_j are zeros of the characteristic equation

$$1 - \sum_{i=1}^{L} a_i z^{-i} = 0 , \qquad (V.8)$$

and λ_j is the residue of G_j. Using the extended ACF, we find the λ_j by solving

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ G_1 & G_2 & \cdots & G_L \\ \vdots & \vdots & & \vdots \\ G_1^{L-1} & G_2^{L-1} & \cdots & G_L^{L-1} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_L \end{pmatrix} = \begin{pmatrix} \gamma(M) \\ \gamma(M+1) \\ \vdots \\ \gamma(M+L-1) \end{pmatrix} . \tag{V.9}$$

The infinite sums in (V.2) can be expressed in terms of one-sided infinite sums, as

$$\sum_{i=-\infty}^{\infty} \gamma(i)\gamma(i+j) = \begin{cases} 2\sum_{i=1}^{\infty} \gamma^2(i) + \gamma^2(0) \; ; \; j = 0 \\ 2\left(\sum_{i=0}^{\infty} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j-1}{2}} \gamma(i)\gamma(j-i)\right) \; ; \; j \text{ odd} \\ 2\left(\sum_{i=0}^{\infty} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j}{2}-1} \gamma(i)\gamma(j-i)\right) + \gamma^2(\tfrac{j}{2}) \; ; \; j \text{ even} \end{cases} \tag{V.10}$$

and the infinite sums can be separated into terms of the ACF directly influenced by the MA parameters and those following the recursive form (V.7), as

$$\sum_{i=0}^{\infty} \gamma(i)\gamma(i+j) = \sum_{i=0}^{M-1} \gamma(i)\gamma(i+j) + \sum_{i=M}^{\infty} \gamma(i)\gamma(i+j)$$

But, from (V.7),

$$\sum_{i=M}^{\infty} \gamma(i)\gamma(i+j) = \sum_{i=M}^{\infty} \sum_{m=1}^{L} \sum_{n=1}^{L} \lambda_m \lambda_n G_m^{\,i-M} G_n^{\,i+j-M} =$$

$$\sum_{m=1}^{L} \sum_{n=1}^{L} \lambda_m \lambda_n \sum_{i=M}^{\infty} G_m^{\,i-M} G_n^{\,i+j-M} = \sum_{m=1}^{L} \sum_{n=1}^{L} \lambda_m \lambda_n \frac{G_n^{\,j}}{1-G_m G_n} \; ; \; j=0,1,2,\ldots, \tag{V.11}$$

where the assumption that all poles are inside the unit circle allows the interchange of summation and guarantees convergence to the closed form result.

Substituting (V.11) into (V.10), we have

$$\sum_{i=-\infty}^{\infty} \gamma(i)\gamma(i+j) = \begin{cases} \gamma^2(0) + 2 \sum_{m=1}^{L} \sum_{n=1}^{L} \frac{\lambda_m \lambda_n}{1-G_m G_n} & ; j=0 \\ 2\left[\sum_{i=0}^{M-1} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j-1}{2}} \gamma(i)\gamma(j-i)\right] + 2 \sum_{m=1}^{L} \sum_{n=1}^{L} \lambda_m \lambda_n \frac{G_n^{\,j}}{1-G_m G_n} & ; j \text{ odd} \\ 2\left[\sum_{i=0}^{M-1} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j}{2}-1} \gamma(i)\gamma(j-i)\right] + \gamma^2(\tfrac{j}{2}) + 2 \sum_{m=1}^{L} \sum_{n=1}^{L} \lambda_m \lambda_n \frac{G_n^{\,j}}{1-G_m G_n} & ; j \text{ even.} \end{cases} \tag{V.12}$$

Finally, we note that (V.2) can be rewritten

$$\lim_{N\to\infty} \operatorname{cov}[c_x(k), c_x(\ell)] = \frac{1}{N} \sum_{i=-\infty}^{\infty} \gamma(i)\gamma(i+\ell-k) + \gamma(i)\gamma(i-\ell-k), \tag{V.13}$$

so that

$$\frac{2}{N}\left\{\gamma^2(0) + 2\sum_{i=1}^{M-1}\gamma^2(i) + 2\sum_{m=1}^{L}\sum_{n=1}^{L}\frac{\lambda_m\lambda_n}{1-G_mG_n}\right\};\ \ell=k=0$$

$$\frac{2}{N}\left\{\sum_{i=1}^{M-1}\gamma^2(i) + \sum_{i=0}^{M-1}\gamma(i)\gamma(i+2\ell) + \right.$$

$$\left.\sum_{i=1}^{\ell-1}\gamma(i)\gamma(2\ell-i) + \sum_{m=1}^{L}\sum_{n=1}^{L}\lambda_m\lambda_n\left(\frac{1+G_n^{2\ell}}{1-G_mG_n}\right)\right\} +$$

$$\frac{1}{N}(\gamma^2(0)+\gamma^2(\ell))\ ;\ \ell = k > 0$$

$$\lim_{N\to\infty}\operatorname{cov}[c_x(k),c_x(\ell)] = \frac{2}{N}\left\{\sum_{i=0}^{M-1}\gamma(i)(\gamma(i+\ell+k)+\gamma(i+\ell-k)) + \right. \tag{V.14}$$

$$\sum_{i=1}^{\frac{\ell+k-1}{2}}\gamma(i)\gamma(\ell+k-i) + \sum_{i=1}^{\frac{\ell-k-1}{2}}\gamma(i)\gamma(\ell-k-i) +$$

$$\left.\sum_{m=1}^{L}\sum_{n=1}^{L}\lambda_m\lambda_n\left(\frac{G_n^{\ell+k}+G_n^{\ell-k}}{1-G_mG_n}\right)\right\}\ ;\ \ell-k = 1,3,5,\ldots$$

$$\frac{2}{N}\left\{\sum_{i=0}^{M-1}\gamma(i)(\gamma(i+\ell+k)+\gamma(i+\ell-k)) + \right.$$

$$\sum_{i=1}^{\frac{\ell+k}{2}-1}\gamma(i)\gamma(\ell+k-i) + \sum_{i=1}^{\frac{\ell-k}{2}-1}\gamma(i)\gamma(\ell-k-i) +$$

$$\left.\sum_{m=1}^{L}\sum_{n=1}^{L}\lambda_m\lambda_n\left(\frac{G_n^{\ell+k}+G_n^{\ell-k}}{1-G_mG_n}\right)\right\} +$$

$$\frac{1}{N}\left(\gamma^2\left(\frac{\ell+k}{2}\right)+\gamma^2\left(\frac{\ell-k}{2}\right)\right)\ ;\ \ell-k = 2,4,6,\ldots,$$

where terms used in the finite sums are obtained using (V.6). Note that the above assumes $\ell \geq k$, $\ell \geq 0$ and $k \geq 0$. This presents no problem in evaluating the covariance matrix, for $c_x(k) = c_x(-k)$ implies that negative values of ℓ and k are not needed, and the toeplitz form of the covariance matrix implies that $\mathrm{cov}[c_x(k), c_x(\ell)] = \mathrm{cov}[c_x(\ell), c_x(k)]$. The case L = 0 is handled by (V.3b) with (V.4) and (V.5). (V.10) is modified to

$$\sum_{i=-\infty}^{\infty} \gamma(i)\gamma(i+j) = \begin{cases} 2\sum_{i=1}^{M} \gamma^2(i) + \gamma^2(0) \; ; \; j = 0 \\ 2\left(\sum_{i=0}^{M} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j-1}{2}} \gamma(i)\gamma(j-i)\right) \; ; \; j \text{ odd} \\ 2\left(\sum_{i=0}^{M} \gamma(i)\gamma(i+j) + \sum_{i=1}^{\frac{j}{2}-1} \gamma(i)\gamma(j-i)\right) + \gamma^2(\tfrac{j}{2}) \; ; \; j \text{ even.} \end{cases} \tag{V.15}$$

Then, using (V.2), we have, for L = 0,

$$\lim_{N\to\infty} \operatorname{cov}[c_x(k),c_x(\ell)] = \begin{cases} \frac{2}{N}\left(2\sum_{i=1}^{M} \gamma^2(i)+\gamma^2(0)\right) ; & \ell=k=0 \\[2ex] \frac{2}{N}\left(\sum_{i=1}^{M} \gamma^2(i) + \sum_{i=0}^{M} \gamma(i)\gamma(i+2\ell) + \sum_{i=1}^{\ell-1} \gamma(i)\gamma(2\ell-i)\right) + \frac{1}{N}(\gamma^2(0)+\gamma^2(\ell)); & \ell=k>0 \\[2ex] \frac{2}{N}\left(\sum_{i=0}^{M} \gamma(i)(\gamma(i+\ell+k)+\gamma(i+\ell-k)) + \sum_{i=1}^{\frac{\ell-k-1}{2}} \gamma(i)\gamma(\ell-k-i) + \sum_{i=1}^{\frac{\ell+k-1}{2}} \gamma(i)\gamma(\ell+k-i)\right) ; & \ell-k=1,3,5,\ldots \\[2ex] \frac{2}{N}\left(\sum_{i=0}^{M} \gamma(i)(\gamma(i+\ell+k)+\gamma(i+\ell-k)) + \sum_{i=1}^{\frac{\ell+k}{2}-1} \gamma(i)\gamma(\ell+k-i) + \sum_{i=1}^{\frac{\ell-k}{2}-1} \gamma(i)\gamma(\ell-k-i)\right) + \frac{1}{N}\left(\gamma^2(\frac{\ell+k}{2})+\gamma^2(\frac{\ell-k}{2})\right) ; & \ell-k=2,4,6,\ldots \end{cases} \tag{V.16}$$

The probability density function (p.d.f.) of the sample ACF of an ARMA process is asymptotically multivariate Gaussian [ ]. If we denote

$$\Gamma^T_{[k_1,k_2]} = [\gamma(k_1),\gamma(k_1+1),\ldots,\gamma(k_2)] \ ; \ k_2 > k_1 \ ,$$

and

$$\Lambda_{[k_1,k_2]} = \lim_{N\to\infty} E(C_{[k_1,k_2]}-\Gamma_{[k_1,k_2]})(C_{[k_1,k_2]}-\Gamma_{[k_1,k_2]})^T \ , \qquad (V.17)$$

then the asymptotic p.d.f. of $C_{[k_1,k_2]}$ is given by

$$\lim_{N\to\infty} f(C_{[k_1,k_2]}|G,P,\sigma_\varepsilon^2) = (2\pi)^{-\frac{k_2-k_1}{2}} |\Lambda_{[k_1,k_2]}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(C_{[k_1,k_2]} - \Gamma_{[k_1,k_2]})^T \Lambda^{-1}_{[k_1,k_2]}(C_{[k_1,k_2]}-\Gamma_{[k_1,k_2]})\right) . \qquad (V.18)$$

where G represents the vector of poles and P the vector of zeros. Fisher's information matrix is given by [45] (denote $\lim_{N\to\infty} f = f_\infty$)

$$I_C(\theta_o) = -E\left(\nabla_{\theta_o}\{[\nabla_{\theta_o} \ln f(C_{[k_1,k_2]}|G,P,\sigma_\varepsilon^2)]^T\}\right) . \qquad (V.19)$$

We take $\theta_o^T = \left(|G_1|,\ldots,|G_{L/2}|,\phi_1,\ldots,\phi_{L/2},|P_1|,\ldots,|P_{M/2}|,\psi_1,\ldots,\psi_{M/2}\right)$, where the assumption that poles and zeros occur in complex conjugate pairs halves the size of the parameter vector. The ϕ_i are angles of the pole pairs, and zeros P_i solve

$$1 - b_1 z^{-1} - \ldots - b_M z^{-M} = 0 \ , \qquad (V.20)$$

each pair having angle ψ_i.

We have been unable to obtain the derivatives in (V.19) due to the complicated nature of the p.d.f., and, even if they were available, it is doubtful that the expectation would yield to a concise solution. Numerical methods are not well developed for multiple integrals, and these have the added complication of being improper, so we consider Monte Carlo methods. The most basic approach is to generate several thousand points $(C_{[k_1,k_2]})_i$ distributed according to (V.18) and to approximate the expectation by a sample mean

$$\hat{I}_C(\theta_o) = \frac{1}{N}\sum_{i=1}^{n}\left\{\nabla_{\theta_o}\{[\nabla_{\theta_o}\ln f_\infty((C_{[k_1,k_2]})_i|G,P,\sigma_\varepsilon^2)]^T\}\right\} =$$

$$\frac{1}{N}\left\{\nabla_{\theta_o}\{[\nabla_{\theta_o}\sum_{i=1}^{n} f_\infty((C_{[k_1,k_2]})_i|G,P,\sigma_\varepsilon^2)]^T\}\right\}, \qquad \text{(V.21)}$$

where N is the number of points generated. The matrix of second partials is computed using standard numerical routines, i.e., we make small perturbations in θ_o. For each new θ_o, we recompute poles $G_i, i=1,\ldots,L$ and zeros $P_i, i=1,\ldots,M$, which occur in conjugate pairs. Then, we obtain the corresponding vectors A and B by solving

$$(1-G_1z^{-1})(1-G_2z^{-1})\ldots(1-G_Lz^{-1}) = 1-a_1z^{-1}-\ldots-a_Lz^{-L}$$

$$(1-P_1z^{-1})(1-P_2z^{-1})\ldots(1-P_Mz^{-1}) = 1-b_1z^{-1}-\ldots-b_Mz^{-M}, \qquad \text{(V.22)}$$

from which we proceed to the calculation of $\gamma(i)$ and eventually $\Lambda_{[k_1,k_2]}$. The sum in (V.21) is evaluated for the several perturbations of θ_o needed in numerical evaluation of the matrix of second partials, using the same set $\{(C_{[k_1,k_2]})_i\}$ each time. If we are

interested in the information with respect only to poles, we use the reduced parameter vector $\theta_o^T = \left(|G_1|,\ldots,|G_{L/2}|,\phi_1,\ldots,\phi_{L/2}\right)$. We then estimate the asymptotic Cramer-Rao bound on the error covariance matrix for $\hat{\theta}_o$ (note that we consider only the class of asymptotically unbiased estimators) as

$$E\left[(\hat{\theta}_o-\theta_o)(\hat{\theta}_o-\theta_o)^T\right] \geq \hat{I}^{-1}(\theta_o) \tag{V.23}$$

where the matrix ordering is

$$R > Q \rightarrow y^T R y > y^T Q y$$

for R and Q square of dimension (say) K and y any vector of dimension K.

Questions regarding the details of the Monte Carlo simulation are under investigation. Next, we will consider the special case of AR time series in white noise of variance σ_n^2. The effect of this noise on the information in $C_{[k_1,k_2]}$ is evaluated by replacing $\gamma(0)$ obtained from (V.3a) by $\gamma'(0) = \gamma(0) + \sigma_n^2$. This will give a quantitative measure of the effect of the model-change phenomenon in the presence of noise. Studies of more general time series and noise will follow.

VI. SPECTRAL ESTIMATION FOR NOISY COMPLEX SINUSOIDS

High resolution spectral estimation has found special popularity in applications involving real or complex sinusoidal signals. Examples of these applications include radar doppler processing, and radar or other sensor array processing for improved angular resolution [27]. A method that has found special appeal in these applications is the maximum entropy method of spectral analysis (MEM) introduced by Burg [8]. However, there have been several disturbing problems with this method when applied to sinusoids, notably, frequency errors depending on the sinusoidal components' phases and line splitting under high order estimates and large signal-to-noise ratios.

We have considered the frequency error problem and have investigated a modification to Burg's original algorithm, which we denote the tapered Burg algorithm. The tapered Burg technique is a direct result of considering a weighted least-squares fit to the parameters of the all-pole model, subject to Levinson's recursion constraint, in place of the usual unweighted least-squares fit. The algorithmic consequence of this approach is the inclusion of an appropriate taper in the calculation of the partial-correlation coefficients in the usual Burg algorithm. Based on the expression for the frequency error in the spectral estimate of a sinusoid, an optimum taper is derived. The performance of this optimum taper is then compared with those of the rectangular (untapered Burg) and Hamming tapers. It appears that this taper makes MEM spectral estimates of sinusoids using Burg's technique more robust, without sacrificing its resolution.

In the popular derivation the MEM coefficients define a predication error filter (PEF) and are chosen so as to minimize the average of the forward and backward prediction residual energies subject to Levinson's recursion constraint. In [8] Burg generalized this derivation by minimizing a weighted average of the residual energies. This was presumably done to allow the analyst to reflect his or her confidence in possibly disjoint data records. For contiguous data, it did not appear that any weighting of the average residuals was needed. Therefore, the popular MEM spectral estimate has been that of Burg's original suggestion of using a rectangular (uniform) taper. Recently, Swingler showed [28], through numerical simulations, that a reduction in the error in the estimated frequency of a real sinusoid is obtained if a Hamming taper is employed in the calculation of Burg's partial-correlation (PARCOR) estimates. This method of tapering is exactly that reported in [8] by Burg.

In this chapter, we first derive an expression for the estimated frequency error of a real sinusoid using Burg's tapered method. This is a generalization of the error expression for the untapered Burg's technique reported by Swingler [29]. We subsequently use the error expression to derive an "optimum" taper. Finally, simulation results comparing the optimum, Hamming and rectangular tapers in obtaining Burg spectral estimates of complex sinusoids in noise are given.

VI.1 Generalized Error Expression

In this section we derive an expression for frequency error of sinusoid based on the tapered Burg technique for a general

weighting function. Let the taper $w_m(t)$ be defined as a function of the continuous variable t for $|t| \leq \frac{N-m}{2}$ and take on the value zero elsewhere. Let the discrete time version be

$$w_{mk} = w_m(k - \frac{N-m}{2}) \quad . \tag{VI.1}$$

As in (6) we assume that the window is normalized and non-negative:

$$\sum_{k=0}^{N-m} w_{mk} = 1 \quad , \quad \text{all } m \ ,$$

$$w_{mk} \geq 0 \quad , \quad \text{all } m,k \ .$$

The non-negativity insures that the magnitude of the partial correlation coefficients (PARCORs) do not exceed unity.

For the data record $x_k, k=0,1,\ldots,N$, the mth PARCOR is given in this tapered Burg method as [8]

$$a_{mm} = -2 \frac{\sum_{k=0}^{N-m} w_{mk} D_{mk} E_{mk}}{\sum_{k=0}^{N-m} w_{mk} (D_{mk}^2 + E_{mk}^2)} \tag{VI.2}$$

where D_{mk} and E_{mk} are given in (VI.6).

For $x_k = \cos(\theta k + \phi)$, the first PARCOR is

$$a_{11} = -\left\{ \cos(\theta) + \frac{\sin^2(\theta) \sum_{k=0}^{N-1} w_{1k} \cos(2\theta k + \theta + 2\phi)}{1 + \cos(\theta) \sum_{k=0}^{N-1} w_{1k} \cos(2\theta k + \theta + 2\phi)} \right\} \tag{VI.3}$$

We choose to express the first PARCOR as

$$\begin{aligned} a_{11} &= -\cos(\theta + \delta) \\ &\simeq -[\cos(\theta) + \delta \sin(\theta)] \ , \ \text{for } \delta << 1 \ . \end{aligned} \tag{VI.4}$$

From (VI.3) we identify δ as

$$\frac{\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi)}{1+\cos(\theta)\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi)} \tag{VI.5}$$

To look at the conditions which make δ small we further assume that the window $w_1(t)$ is even and thus its Fourier transform $W_1(\omega)$ is real. The summation in (VI.5) then becomes

$$\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi) = \cos(N\theta+2\phi)\sum_{n=-\infty}^{\infty}(-1)^{n(N-1)}W_1(2\pi n+2\phi) \tag{VI.6}$$

Since $W_1(\omega)$ is bandlimited on the order of $\frac{2\pi}{N-1}$ r/s we can approximate the summation when $|\theta| < \pi$ and $N >> 1$ as

$$\begin{aligned}\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi) &\simeq \cos(N\theta+2\phi)W_1(2\theta)\\ &\simeq 0\ ,\ \frac{\pi}{N-1} < |\theta| < \pi - \frac{\pi}{N-1}\end{aligned} \tag{VI.7}$$

From (VI.6) and the constraints on w_{mk} we have that $W_1(0) \simeq 1$. Thus we see that the assumption $\delta << 1$ is valid when $|\theta| < \pi$ and $N >> 1$.

The second PARCOR is exactly given by

$$a_{22} = -\frac{A}{B}\left(\frac{1+\frac{B}{A}\sum_{k=0}^{N-2} w_{2k}\cos(2\theta k+2\theta+2\phi)}{1+\frac{A}{B}\sum_{k=0}^{N-2} w_{2k}\cos(2\theta k+2\theta+2\phi)}\right) \tag{VI.8}$$

where

$$A = a_{11}^2 + 2a_{11}\cos(\theta) + \cos(2\theta)$$

$$B = a_{11}^2 + 2a_{11}\cos(\theta) + 1$$

Using the approximation for a_{11} results in

$$A = -(1-\delta^2)\sin(\theta) \ , \ B = (1-\delta^2)\sin(\theta) \ ,$$

and thus

$$a_{22} \simeq 1 \ , \ \delta << 1 \ .$$

The Levinson recursion gives $a_{21} = a_{11} + a_{22}a_{11} \simeq 2a_{11}$ and the resulting PEF is then $(1, -2\cos(\theta-\delta), 1)$. Comparing this to the ideal PEF $(1, -2\cos(\theta), 1)$, we identify the frequency error as

$$\Delta f = \frac{\delta}{2\pi}$$

$$= \frac{1}{2\pi}\sin(\theta)\frac{\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi)}{1+\cos(\theta)\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi)} \qquad \text{(VI.9)}$$

which may be approximated by

$$\Delta f \simeq \frac{1}{2\pi}\sin(\theta)\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi), \qquad \text{(VI.10(a))}$$

$$\frac{\pi}{N-1} < |\theta| < \pi - \frac{\pi}{N-1}$$

$$\simeq \frac{1}{2\pi}\sin(\theta)\cos(N\theta+2\phi)W_1(2\theta) \ . \qquad \text{(VI.10(b))}$$

For uniform weighting, $w_{1k} = \frac{1}{N}$, we have

$$\sum_{k=0}^{N-1} w_{1k}\cos(2\theta k+\theta+2\phi) = \frac{1}{N}\cos(N\theta+2\phi)\frac{\sin(N\theta)}{\sin(\theta)}$$

and (10a) reduces to Swingler's expression [29].

VI.2 An Optimum Taper

If we consider the phase ϕ as a random variable uniformly distributed on $[-\pi,\pi)$ then the mean value of the frequency error is zero, using (10a). The variance of this frequency error is then

$$\mathrm{var}(\Delta f) = \frac{1}{8\pi^2}\sin^2(\theta)\sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} w_{1k}w_{1\ell}\cos(2\theta(k-\ell)). \tag{VI.11}$$

As a criterion for selecting a taper we use the average frequency error variance:

$$\langle \mathrm{var}(\Delta f)\rangle = \frac{1}{\pi}\int_0^{\pi} \mathrm{var}(\Delta f)\,d\theta$$

$$= \frac{1}{8\pi^3}\sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} w_{1k}w_{1\ell}\left(\frac{1}{2}\delta_{k-\ell} - \frac{1}{4}\delta_{1-k+\ell} - \frac{1}{4}\delta_{1+k-\ell}\right), \tag{VI.12}$$

where the subscripted delta is the digital (Kronecker) impulse. With the normalization constraint introduced using the Lagrange multiplier λ the resulting optimum taper is given by the solution of

$$\underline{C}\,\underline{W}_1 = \underline{\lambda}\ ,\ c_{ii} = 2, c_{ij}\begin{cases}-1\ , & j = i\pm 1\\ 0\ , & \text{otherwise}\end{cases},\ \lambda_i=\lambda \tag{VI.13}$$

The general system of equations with a tri-diagonal coefficient matrix has a known recursive solution related to the LU decomposition of the coefficient matrix (see for example [30]). For the system of equations with the special coefficient matrix in (VI.13), we have derived closed form expressions for the taper coefficients (Appendix B). These are given for the m-th order PARCOR by:

$$w_{mk} = \frac{6(k+1)(N-m-k+1)}{(N-m+1)(N-m+2)(N-m+3)}\ ,\ k=0,\ldots,N-m \tag{VI.14}$$

Thus, this optimum taper is parabolic in form, it is even, positive and has a maximum at $k = \frac{N-m+1}{2}$. Furthermore, knowing w_{mo} and w_{m1} from (VI.14), one can generate the remaining coefficients recursively from:

$$w_{mk} = 2w_{m(k-1)} - w_{m(k-2)} - \lambda \qquad \text{(VI.15)}$$

VI.3 Simulation Results

The effect of tapering was numerically investigated by generating spatial samples of complex sinusoids in noise for various, phase and angle (frequency) combinations. For brevity of space, however, only the results of a very few representative simulations are included in this report. The method designations on the plots are: Burg for untapered Burg method, WBurgH and WBurgO for tapered Burg technique using the Hamming taper and the optimum taper respectively. The Hamming taper is used for comparison with the simulations in [28].

Figure (VI.1) shows three tapered Burg spectral estimates of two complex sinusoids at -30° and -45° off broadside with zero phases each, in complex white noise. The signal-to-noise ratio is 15 dB where $\text{SNR(dB)} = 10 \log(\frac{\text{Amplitude}}{\sigma_n^2})$. It is obvious in this example that WBurgO has the smallest frequency errors as well as the highest resolution. This has been the case in the great majority of the examples that we have run thus far. Except for a few cases the bias of WBurg optimum has been lower than Burg. The resolution of WBurg Hamming, however, was found to be consistently poorer than Burg and WBurgO. Figure (VI.2) shows the spectral estimate of a real sinusoid with a large SNR. It can be seen that

WBurg optimum taper has reduced the frequency error and prevented peak splitting which is shown by the untapered Burg method of order 4. This again was accomplished without sacrificing resolution.

Figure (VI.3) is a plot of the maximum frequency error for each method for a unit amplitude real sinusoid at angle (frequency) of $\theta = \pi/5$ as a function of the number of data samples. The maximum frequency errors were found by solving for the pole locations over the range of values for the sinusoid phase ϕ. The noise variance is $\sigma_n^2 = 0.025$. It should be noted that the optimum taper was derived on the basis of minimum <u>average frequency variance</u>. This simulation shows comparative error reduction for the <u>maximum frequency error</u> only for $\theta = \pi/5$ between the optimum and Hamming tapers. As was pointed out earlier, however, all of our simulations have shown better resolution for the optimum taper compared to the Hamming one. It is also interesting to note that the main reduction in frequency error is apparently obtained by tapering the first order residual energy only. This has also been borne out by the majority of our simulations.

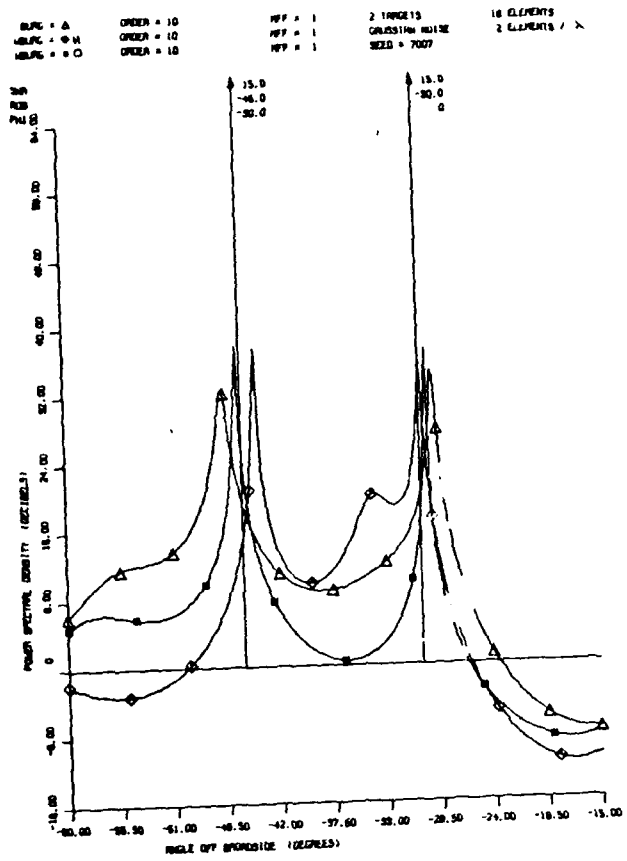


FIGURE VI.1 Spectral estimates of two complex exponentials in white noise.

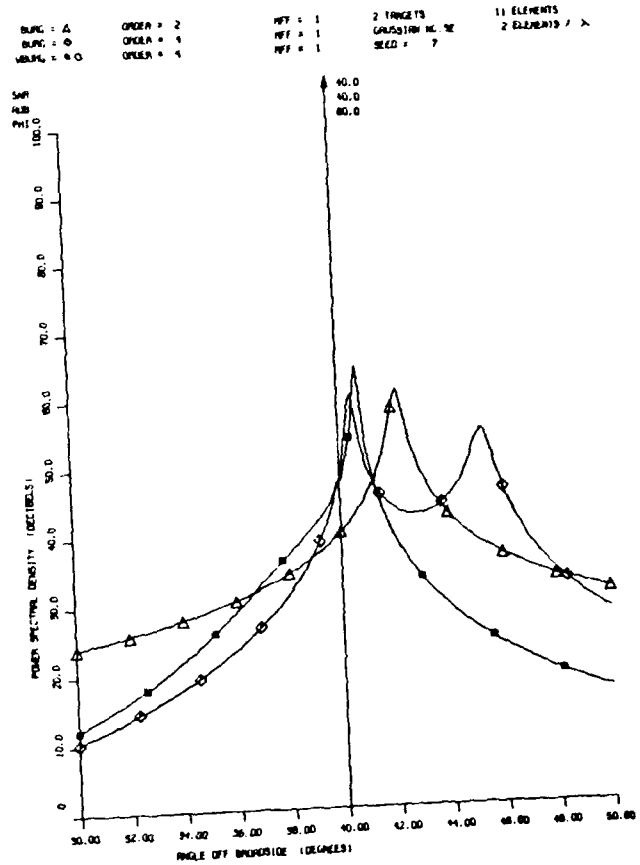


FIGURE VI.2 Spectral estimates of two complex conjugate exponentials (one real sinusoid) in white noise.

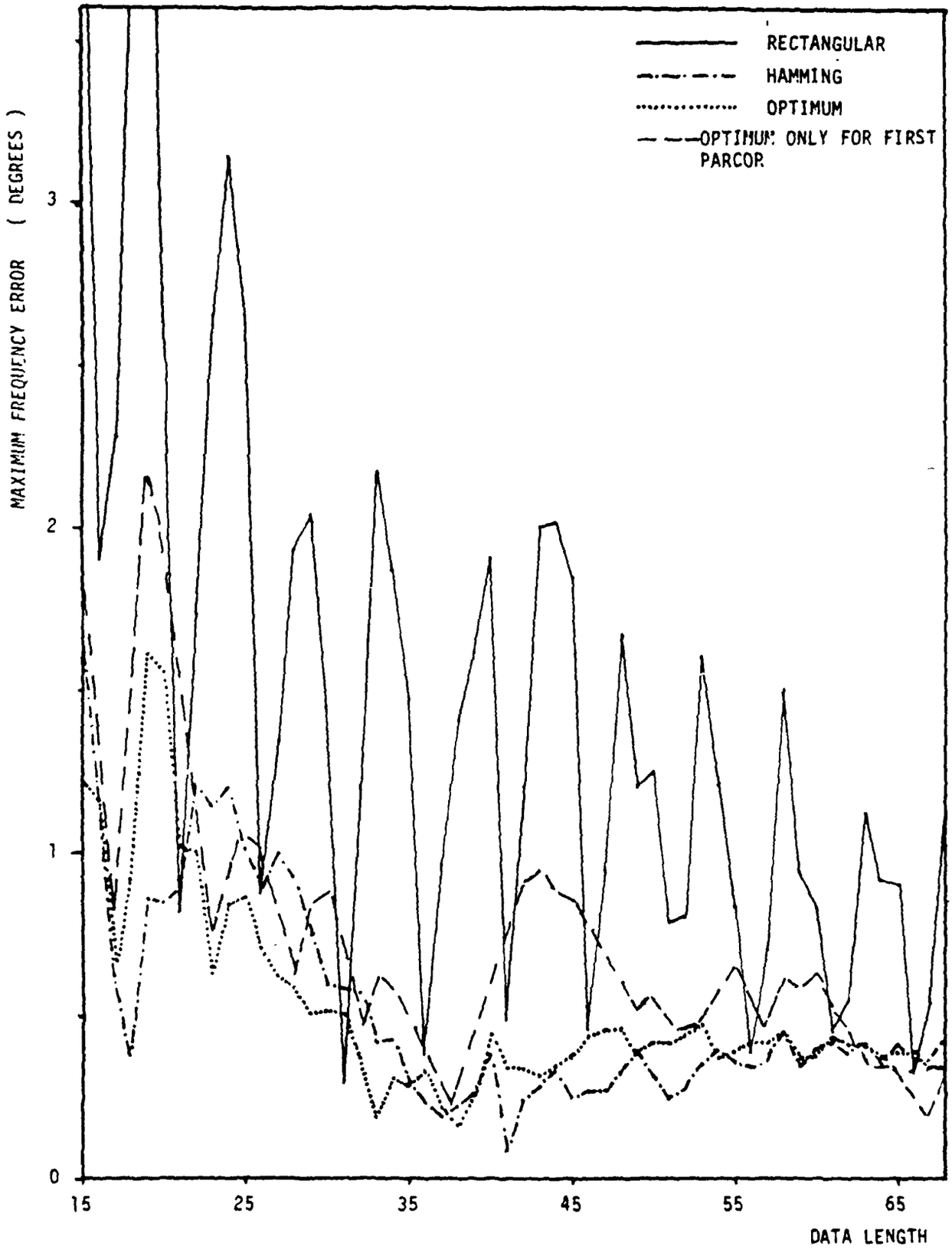


FIGURE VI.3 Maximum errors in the estimated frequency of a real sinusoid as a function of data length.

VII. SUMMARY AND CONCLUSIONS

Data-adaptive spectral estimation methods have found widespread application in recent years, notably in radar signal processing. The work summarized in this report was concerned with two such methods: the autoregressive (AR) and the autoregressive-moving average (ARMA) spectral estimators. Our main concern in this investigation was the resolving capabilities of the different models in the presence of noise and the statistical properties of some new spectral estimators.

Since the AR methods are by far the most popular, we postulated an AR noiseless signal model. The resulting appropriate model in the presence of noise was then shown to be ARMA. To investigate the degradation in resolution in the spectrum of noisy signals, we also used a parallel resonator model for the signal. This model was also shown to be equivalent to an ARMA one. The resonator model, however, made it possible to evaluate the loss in resolution as a function of signal-to-noise ratio and signal parameters.

Since an ARMA model may be approximated by a high-order AR one, attention was next focussed on the development of a robust method for identifying the order of an AR model. A method closely related to that of Akaike's information criterion was developed. This method proved to be more stable than the minimum AIC.

ARMA spectral estimation was treated extensively in this work. The generality of the model was important from two distinct points of view: i) as the appropriate model for noisy AR signals, thus to improve resolution; ii) as a resonable model for signals with

spectra, narrow or wide, possessing deep nulls and/or sharp roll-offs. Because of the computational complexities of the optimum (maximum likelihood) method of estimating the ARMA parameters, several sub-optimum, computationally efficient techniques were devised. These methods behaved reasonably well for moderate to large data samples, but were inferior to the optimum one, as expected, in terms of statistical efficiency. The efficiency of various classes of ARMA spectral estimators are currently under further investigation by Monte Carlo simulations and will be reported on in the future.

The maximum entropy method (MEM) of spectral estimation, using Burg's technique, has been very popular in estimating the spectra of sinusoidal type signals. It has been known, however, that the accuracy of such spectral estimates is significantly influenced by the phase of the sinusoids. Based on some recent results, we developed an optimum taper for the residual energies in the Burg MEM technique. It was shown that the use of such a taper substantially reduced the sensitivity of the spectral peaks to the sinusoids' phases and markedly reduced the occurrence of line-splitting in the usual MEM estimates. It is remarkable that these improvements were made without sacrifice in the resolution of the spectral estimates. Work is continuing to more fully explain the effect of such a taper and its ramifications in Burg-type parameter estimation techniques.

APPENDIX A

Equation (II.25) is

$$a'_{m2} = \frac{r_{y_m}(0)r_{y_m}(2) - r^2_{y_m}(1)}{r^2_{y_m}(0) - r^2_{y_m}(1)} \tag{A.1}$$

$$= -\rho'^2_m$$

Now substitute for $r_{y_m}(0)$, $r_{y_m}(2)$ and $r_{y_m}(1)$ in A.1,

$$\frac{(a^2_{mo}\sigma^2_u\alpha\cos\phi_m+\sigma^2_{w'})(a^2_{mo}\sigma^2_u\alpha\rho^2_m\cos(2w_m-\phi_m))-(a^2_{mo}\sigma^2_u\alpha\rho_m\cos(w_m-\phi_m))^2}{(a^2_{mo}\sigma^2_u\alpha\cos\phi_m+\sigma^2_{w'})^2-(a^2_{mo}\sigma^2_u\alpha\rho_m\cos(w_m-\phi_m))^2}$$

$$= -\rho'^2_m \tag{A.2}$$

Expanding the numerator and denominator of equation A.2, we get,

$$\frac{(a^2_{mo}\sigma^2_u)^2\alpha^2\rho^2_m\cos\phi\cos(2w_m-\phi_m)+a^2_{mo}\sigma^2_u\sigma^2_{w'}\alpha\rho^2_m\cos(2w_m-\phi_m) - (a^2_{mo}\sigma^2_u)^2\alpha^2\rho^2_m\cos^2(w_m-\phi_m)}{(a^2_{mo}\sigma^2_u)^2\alpha^2\cos^2\phi_m+\sigma^4_{w'}+2a^2_{mo}\sigma^2_u\sigma^2_{w'}\alpha\cos\phi_m - (a^2_{mo}\sigma^2_n)^2\alpha^2\rho^2_m\cos^2(w_m-\phi_m)}$$

$$= -\rho'^2_m \tag{A.3}$$

Define the signal-to-noise ratio ξ as ,

$$\xi = \frac{r_{y_m}(0) - \sigma_{w'}^2}{\sigma_{w'}^2}$$

$$= \frac{r_{y_m}(0)}{\sigma_{w'}^2} - 1$$

$$= \gamma - 1 \tag{A.4}$$

where,

$$r_{y_m}(0) = a_{mo}^2 \sigma_u^2 \alpha \cos\phi_m$$

Dividing the numerator and denominator of A.3 by $\sigma_{w'}^4$, and using A.4, we get,

$$\frac{\left(\gamma^2 \frac{\cos(2w_m-\phi_m)}{\cos\phi_m} + \gamma \frac{\cos(2w_m-\phi_m)}{\cos\phi_m} - \gamma^2 \frac{\cos^2(w_m-\phi_m)}{\cos^2\phi_m}\right)\rho_m^2}{\gamma^2 + 1 + 2\gamma - \gamma^2 \frac{\rho_m^2\cos^2(w_m-\phi_m)}{\cos^2\phi_m}}$$

$$= -\rho'^2_m \tag{A.5}$$

Rearranging terms in A.5, we have

$$\frac{\gamma^2\left(\cos\phi_m\cos(2w_m-\phi_m)-\cos^2(w_m-\phi_m)\right)+\gamma\cos\phi_m\cos(2w_m-\phi_m)}{\gamma\left(\cos^2\phi_m-p_m^2\cos^2(w_m-\varphi_m)\right)+2\gamma\cos^2\phi_m+\cos^2\phi_m}$$

$$= -\frac{\rho'^2_m}{\rho_m^2} \tag{A.6}$$

As $\gamma \to \infty$, $-\frac{\rho'^2_m}{\rho^2_m} \to 1$ and the left-hand side of equation A.6 is,

$$\frac{\cos\phi_m\cos(2w_m-\phi_m)-\cos^2(w_m-\phi_m)}{\cos^2\phi_m-\rho_m^2\cos^2(w_m-\phi_m)} = -1 \tag{A.7}$$

Solving for $\rho_m^2\cos^2(w_m-\phi_m)$ in A.7 we get,

$$\rho_m^2\cos^2(w_m-\phi_m)=\cos\phi_m\cos(2w_m-\phi_m)-\cos^2(w_m-\phi_m)$$
$$+\cos^2\phi_m \tag{A.8}$$

Substitute A.8 in A.6 to get,

$$\frac{\gamma^2\left[\cos\phi_m\cos(2w_m-\phi_m)-\cos^2(w_m-\phi_m)\right]+\gamma\cos\phi_m\cos(2w_m-\phi_m)}{\gamma^2\left[-\cos\phi_m\cos(2w_m-\phi_m)+\cos^2(w_m-\phi_m)\right]+2\gamma\cos^2\phi_m+\cos^2\phi_m}$$
$$= -\frac{\rho'^2_m}{\rho_m^2} \tag{A.9}$$

But the signal-to-noise ratio is (equation A.4)

$$\xi = \gamma-1$$

A.7 then becomes,

$$\frac{(\xi+1)^2\left[\cos\phi_m\cos(2w_m-\phi_m)-\cos^2(w_m-\phi_m)\right]+(\xi+1)\cos\phi_m\cos(2w_m-\phi_m)}{(\xi+1)^2\left[-\cos\phi_m\cos(2w_m-\phi_m)+\cos^2(w_m-\phi_m)\right]+2(\xi+1)\cos^2\phi_m+\cos^2\phi_m}$$
$$= -\frac{\rho'^2_m}{\rho_m^2} \tag{A.10}$$

Equation A.10 simplifies to,

$$\frac{\xi^2\left[\cos^2(w_m-\phi_m)-\cos\phi_m\cos(2w_m-\phi_m)\right]-\xi\left[3\cos\phi_m\cos(2w_m-\phi_m)-2\cos^2(w_m-\phi_m)\right]-2\cos\phi_m\cos(2w_m-\phi_m)+\cos^2(w_m-\phi_m)}{\xi^2\left[\cos^2(w_m-\phi_m)-\cos\phi_m\cos(2w_m-\phi_m)\right]+2\xi\left[\cos^2\phi_m-\cos\phi_m\cos(2w_m-\phi_m)+\cos^2(w_m-\phi_m)\right]+3\cos^2\phi_m-\cos\phi_m\cos(2w_m-\phi_m)+\cos^2(w_m-\phi_m)}$$

$$= \frac{\sigma'^2_m}{\sigma^2_m} \tag{A.11}$$

APPENDIX B

In this appendix we derive a closed form expression for the solution of the following equations:

$$\begin{pmatrix} 2 & -1 & 0 & \dots & 0 & \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ 0 & \dots & 0 & -1 & 2 & \end{pmatrix} \begin{pmatrix} w_1 \\ \\ \\ w_N \end{pmatrix} = \begin{pmatrix} \lambda \\ \cdot \\ \cdot \\ \cdot \\ \lambda \end{pmatrix} \tag{B.1}$$

where λ is the normalization factor for $\underline{W}$.

A recursive algorithm for a set of linear equations with a general tridiagonal coefficient matrix is given in [30]. We use the notation used in this reference to derive our closed form solution.

The solution for $w_1 = w_N$ is given by

$$w_N = w_1 = \frac{g_N}{\alpha_N} \tag{B.2}$$

where according to [30]

$$\begin{aligned} g_k &= \lambda - \frac{(k-1)}{k} g_{k-1} \ , \ k=1, \ \dots \ ,N \\ \alpha_k &= 2+\beta_k \ , \ k=2, \ \dots \ ,N \\ \alpha_1 &= 2 \\ \beta_k &= - \frac{1}{\alpha_{k-1}} \end{aligned} \tag{B.3}$$

We first derive an expression for β_k. Let $\beta_k = -\nu_k/\delta_k$, where ν_k and δ_k are integers.

From (B.3)

$$\beta_k = - \frac{1}{2 + \beta_{k-1}} \ , \ k=2, \ \dots \ ,N \tag{B.4}$$

Then

$$\frac{\nu_k}{\delta_k} = \frac{1}{2 + \frac{\nu_{k-1}}{\delta_{k-1}}} = \frac{\delta_{k-1}}{2\delta_{k-1} - \nu_{k-1}} \tag{B.5}$$

Now let $\delta_{k-1} = \nu_{k-1} + m_{k-1}$ then

$$\frac{\nu_k}{\delta_k} = \frac{\delta_{k-1}}{\delta_{k-1} + m_{k-1}} \tag{B.6}$$

But we have $\nu_2 = 1$, $\delta_2 = 2$ and $m_2 = 1$ and from (B.6) it is obvious that m_k = constant. Therefore,

$$\frac{\nu_k}{\delta_k} = \frac{\delta_{k-1}}{\delta_{k-1+1}}$$

or with $\nu_2 = 1$, $\delta_2 = 2$

$$\beta_k = \frac{-(k-1)}{k} \tag{B.7}$$

α_k now simply follows as

$$\alpha_k = 2+\beta_k = \frac{k+1}{k} \text{ , } k=1, \ldots ,N \tag{B.8}$$

Let $h_k = kg_k$ and substitute for q_k in (B.3). h_k then satisfies the difference equation

$$h_k - h_{k-1} = k\lambda \text{ , } h_o = 0 \tag{B.9}$$

Z-transform (B.9) to get

$$H(z) = \frac{z}{z-1} \cdot \frac{z\lambda}{(z-1)^2} \tag{B.10}$$

resulting in

$$h_k = (\text{step})*(\text{ramp})\lambda = \lambda \sum_{j=1}^{k} j$$
$$= \frac{k(k+1)}{2}\lambda \tag{B.11}$$

Therefore

$$g_k = \frac{k+1}{2}\lambda \tag{B.12}$$

Substitute in (B.2) to get

$$w_N = w_1 = \frac{N\lambda}{2} \tag{B.13}$$

Using (B.13) as the initial condition we can now solve for other w_i's, noting that

$$w_k - 2\,w_{k-1} + w_{k-2} = -\lambda\ ,\ k=2,\ \ldots\ ,N \tag{B.14}$$
$$w_o = 0\ ,\ w_1 = \frac{N\lambda}{2}$$

Z-transform of (B.14) results in

$$W(z) = \frac{-z^3\lambda}{(z-1)^3} + \frac{(N+2)z\lambda}{2(z-1)^2} + \frac{\lambda z^2}{(z-1)^2} \tag{B.15}$$

Taking the inverse transform of (B.15) gives

$$w_k = -\frac{\lambda(k+1)(k+2)}{2} + \frac{(N+2)k\lambda}{2} + (k+1)\lambda$$

or

$$w_k = \frac{k(N+1-k)}{2}\lambda \tag{B.16}$$

The normalization factor λ can now be found as

$$\lambda = \left(\sum_{k=1}^{N} w_k\right)^{-1} = \frac{12}{N(N+1)(N+2)} \tag{B.17}$$

Alternatively,

$$\lambda_{N+1} = \lambda_N \cdot \frac{N}{N+3} \tag{B.18}$$

REFERENCES

1. R. T. Lacoss, "Data adaptive spectral analysis methods", Geophysics, vol. 36, pp. 661-675, Aug. 1971.

2. M. Kaveh and G. R. Cooper, "An empirical investigation of the properties of the autoregressive spectral estimator", IEEE Trans. Info. Theory, vol. IT-22, pp. 313-323, May, 1976.

3. R. E. Kromer, "Asymptotic properties of the autoregressive spectral estimator", Ph.D. dissertation, Stanford Univ., Stanford, Calif., Dec. 1970.

4. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra from the Point of View of Communications Engineering, New York: Dover, 1959.

5. K. N. Berk, "Consistent autoregressive spectral estimates", ANN. Statist., vol. 2, pp. 489-502, May, 1974.

6. A. B. Baggeroer, "Confidence intervals for regression (MEM) spectral estimates", IEEE Trans. Info. Theory, vol. IT-22, pp. 534-545.

7. G. E. P. Box and G. M. Jenkins, Time Series, Forecasting and Control, San Francisco: Holden-Day, 1970.

8. J. P. Burg, "Maximum entropy spectral analysis", Ph.D. Thesis, Stanford University, Palo Alto, CA, 1975.

9. R. McAulay, "Maximum-likelihood spectral estimation using state-variable techniques", Proceedings of First RADC Spectrum Estimation Workshop, Rome, NY, 1978.

10. J. Haddad, "Spectral estimation of noisy signals using a parallel resonator model and state-variable techniques", M.S. paper, University of Minnesota, Minneapolis, MN, 1980.

11. S. Kay, "Autoregressive spectral analysis of sonar signals", Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, 1979.

12. G. M. Jenkins and D. G. Watts, "Spectral analysis and its applications", San Francisco, Holden-Day, 1968.

13. H. Akaike, "A new look at statistical model identification", IEEE Trans. on Automatic Control, vol. AC-19, no. 6, December 1974.

14. H. Akaike, "A bayesian extension of the minimum AIC procedure of autoregressive model fitting", Research Memo No. 126, The Institute of Statistical Mathematics, November 1977.

15. E. Parzen, "Multiple time series: determining the order of approximating autoregressive schemes", Tech. Report No. 23, July, SUNY Buffalo Dept. Computer Science, 1975.

16. R. J. Bhansali and D. Y. Downham, "Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion", Biometrika, vol. 64, 1977.

17. M. Kaveh, "A modified Akaike information criterion", Proceedings of the 17th CDC, January 1979.

18. B. W. Lindgren, Statistical Theory, 3rd ed., Macmillan, New York, 1976.

19. S. Bruzzone and M. Kaveh, "On some suboptimum ARMA spectral estimators", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, pp. 753-755, December 1980.

20. J. A. Cadzow, "ARMA spectral estimation, a model equation error procedure", presented at the 1980 Int. Conf. Acoust., Speech, Signal Processing, Denver, CO, April 14-16, 1980.

21. M. Kaveh, "High resolution spectral esimtation for noisy signals", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 286-287, June 1979.

22. J. F. Kinkel, J. Perl, L. L. Scharf, and A. R. Stubberud, "A note on covariance-invariant digital filter design and autoregressive moving average spectrum analysis", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, April 1979.

23. A. Papoulis, Probability, Random Variables, and Stochastic Processes, New York: McGraw-Hill.

24. G. E. Shilov and R. A. Silverman, translator, Linear Algebra, New York: Dover, 1977.

25. E. Parzen, "An approach to time series analysis", Ann. Math. Statist., vol. 32, pp. 951-986, Dec. 1961.

26. M. J. D. Powell, "Restart procedures for the conjugate gradient method", Math. Programming, vol. 12, pp. 241-254, 1977.

27. J. Evans, "Aperture sampling techniques for precision direction funding", IEEE Trans. Aerospace and Electronic Systems, vol. 15, Nov. 1979.

28. D. N. Swingler, "A modified Burg algorithm for maximum entropy spectral analysis", Proc. IEEE, vol. 67, no. 9, Sept. 1979.

29. D. N. Swingler, "Frequency errors in MEM processing", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, no. 2, April 1980.

30. G. Dahlquist and A. Björk, Numerical Methods, Prentice-Hall, 1974.

31. M. Arato, "On the sufficient statistics for stationary Gaussian random processes", Theo. Prob. Appl., vol. 6, no. 2, pp. 199-201, 1961.

32. S. A. Tretter and K. Steiglitz, "Power spectrum identification in terms of rational models", IEEE Trans. Automat. Contr., pp. 185-188, April 1967.

33. E. J. Hannan, "The estimation of mixed moving average autoregressive systems", Biometrika, 56, 3, pp. 579-593, 1969.

34. H. Akaike, "Maximum likelihood identification of Gaussian autoregressive moving-average models", Biometrika 60, 2, pp. 255-265, 1973.

35. I. S. Konvalinka and M. R. Matausek, "Simultaneous estimation of poles and zeros in speech analysis and ITIF-iterative inverse filtering algorithm", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, no. 5, pp. 485-492, Oct. 1979.

36. A. M. Walker, "Large-sample estimation of parameters for autoregressive processes with moving-average residuals", Biometrika, 49, pp. 117-131, 1962.

37. T. C. Hsia and D. A. Landgrebe, "On a method for estimating power spectra", IEEE Trans. Instrumentation and Measurement, vol. IM-16, no. 3, pp. 255-257, Sept. 1967.

38. D. Graupe, D. J. Krause, and J. B. Moore, "Identification of autogressive moving-average parameters of time series", IEEE Trans. Automat. Contr., pp. 104-107, Feb. 1975.

39. H. Sakai and M. Arase, "Recursive parameter estimation of an autoregressive process disturbed by white noise", Int. J. Control, vol. 30, no. 6, pp. 949-966, 1979.

40. E. H. Satorius and S. T. Alexander, "High resolution spectral analysis of sinusoids in correlated noise", 1978 IEEE ICASSP, Tulsa, OK, April 10-12, 1978.

41. S. M. Kay, "A new ARMA spectral estimator", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, pp. 585-588, Oct. 1980.

42. W. Gersch, "Estimation of the autoregressive parameters of a mixed autoregressive moving average time series", IEEE Trans. Automat. Contr., pp. 583-588, Oct. 1970.

43. H. Sakai and H. Tokumaru, "Statistical analysis of a spectral estimator for ARMA processes", IEEE Trans. Automat. Contr., vol. AC-25, pp. 122-124, Feb. 1980.

44. S. Bruzzone, "Statistical analysis of some suboptimum ARMA spectral estimators", Ph.D. dissertation, Univ. of Minnesota, Minneapolis, Minn., June 1981.

45. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I., New York: John Wiley and Sons, Inc., 1968.

DAT
FILM